Testing procedures and acquisition systems for contact sensor¿based vocal monitoring devices

ScuDo

Scuola di Dottorato ~ Doctoral School

WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Metrology (30^{th}cycle)

Testing procedures and acquisition systems for contact sensor–based vocal monitoring devices

By

Federico Casassa

******

Supervisor(s):

Alberto Vallan

Arianna Astolfi

Doctoral Examination Committee:

Politecnico di Torino

2019

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Federico Casassa
2019

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Alla mia compagna Elisa,
che mi ha sostenuto e spronato fino alla fine,
che ha creduto in me
anche quando io stesso ci credevo poco.

Alla mia famiglia,
che nonostante tutto non ha mai smesso di darmi fiducia e comprensione.

Abstract

The goal of the research activity described in this dissertation was to find new solutions for vocal monitoring, specifically concerning the improvement of the data acquisition device and the development of test systems and test procedures to estimate the uncertainty related to the measurement arrangements.
Voice disorders, from simple vocal dysphonia to vocal fold nodules, could affect those professionals who use voice in a sustained way for long intervals, e.g. actors, singers, call-center operators. Monitoring the vocal activity of these people during a normal working day is useful to identify incorrect vocal behaviour, and possible voice disorders.
The device used for this procedure (long-term voice monitoring) is called "vocal dosimeter". It allows to record vocal data while the monitored subject performs his/her normal activities.

The first part of the work deals about the improvement of the Voicecare vocal dosimeter, developed at the Politecnico di Torino, DET Department, in collaboration with TEBE research Group. The system is capable to record the vibration signal at the base of the neck, which is directly connected to the phonation. The acquired signal, post-processed, could provide information about the voice use and possible vocal disorders of the monitored subjects. One of the goals was to lower the uncertainty related to the instrument by searching for a better contact sensor, so three of them were selected to be tested with the one already used on the dosimeter. For this purpose in vivo test procedures was developed, useful to select the best contact sensor, but also to obtain information about the uncertainty related to the sensor itself. They were used to acquire different vocal tasks performed by various test subjects in a semianechoic room. While, the tests gave some good results but exposed that there was more work to do in evaluating the frequency response of each sensor used for this particular task. Due to the biological variability, this kind of test performed on

human subjects does not provide a strong reference for a correct sensor comparison and evaluation.
Alongside, enhancements were made on the data logger section of the vocal dosimeter. The existing micro-controlled board used on the original one was replaced with a more powerful one, and the increased computing power has allowed to real-time estimate some vocal parameters, which are usually obtained by postprocessing the data. Also, a programmable gain amplifier was implemented on the input channels: the automation of the gain has helped to reduce the saturation and to improve the acquisition quality.
The system was tested in order to define the uncertainty in the parameter estimation, and to tune the gain automation algorithm.

In the second part of the dissertation, the issues emerged after the tests for the selection of a new contact sensor has been dealt. The in vivo tests suffered from lack of repeatability due to the biological variability and to the absence of a strong reference. For this purpose, a phonatory system simulator has been developed and tested, with the aim of obtaining an apparatus that provides a stable reference to test and evaluate contact sensors to be used for voice monitoring.
Such apparatus must provide a stable relation between a vibration on a skin-like material and an acoustic output. The simulator has been characterized and tested in terms of measurement repeatability and agreement with the true phonatory system response, and a method to evaluate the sensors load effect has been proposed.

The tests on human subject helped to resolve some problems related to the data acquisition in vocal monitoring (vocal task to be performed, contact sensor attachment method) and the results obtained were used to better calibrate the acquisition system.
The work on the data logger improved the acquired signal quality (less saturated frames, more data to process, real-time parameter estimation with better accuracy) and the development of the phonatory system simulator has provided a test system for the contact sensors and for the acquisition system to be used for vocal monitoring. The simulator has proved to be a stable reference on which repeatable measurements can be performed, and it has already been used to improve the calibration of the contact sensor channel of the Voicecare, to obtain the uncertainty related to the defintion of the calibration function.

Contents

List of Figures

List of Tables

Preface

The work discussed in this dissertation is part of a bigger project that deals with vocal monitoring.
The activities of the research group involved in this project cover the development of instruments and techniques for recording the voice activity of human subjects [1,2], their metrological characterization [3-6], the analysis of the collected data and the study of new parameters, extracted from the voice activity recordings, which can be useful in vocal pathology assessment and prevention. The applications of the results are various: the prevention of the vocal pathologies in teachers [7-9], the ambulatory assessment of vocal diseases and the voice quality monitoring in singers.

The group worked on the development of a vocal dosimeter, the Voicecare. It is a data logger based on a microcontroller board (Arduino Uno), which acquires the vocal data by means of two different transducers: a contact sensor and a normal air microphone.
The phonation is the results of the air that flows through the vibrating vocal folds; this creates a vibration on the skin of the neck and a pressure wave which travels through the vocal tract. The vocal tract filters the pressure wave according to its resonances and "became" the actual voice, the phonation acoustic emission. The skin vibration is sensed by means of the contact sensor, and the phonation acoustic emission is acquired by means of an air microphone.

The Voicecare is buildt with low-cost components and it is easy to use in different situations. Its working principle will be explained more in depth in the first chapter. The prototype has been tested in order to quantify the uncertainty related to the calibration procedure [1-6], and it has been used to acquire data for several studies [7-9]. These studies point out that the system needs some modification, in order to

lower the uncertainty due to the device itself and to obtain some parameters in real time. The Voicecare was also compared with three other vocal dosimeters [4] in order to quantify the measurement accuracy.
The work presented in this dissertation deals with the improvement of the acquisition system and the development of tools and procedures to define some of its metrological properties.

The first chapter is focused on the improvement of the vocal dosimeter itself, by means of the selection of the contact sensor and the improvement of the data logger electronic parts.
Several types of contact sensor are available on the market. They are based on different technologies since they are designed for different applications. The one used in the development prototype is a laringophone originally designed for helicopter pilots, modified in order to be used with the Voicecare. Some other contact sensors were selected to be tested in a voice monitoring framework, in order to identify the best sensor to be used with the Voicecare. For this purpose, test procedures has been designed and used on the selected sensors. These procedures are basically vocal and physical tasks performed by human subjects wearing the selected sensors. The tasks are designed to simulate different conditions and situations in which the voice monitoring could take place, such as lesson, a call center working routine, an ambulatory medical examination. These tests, which are based on those performed in [4], highlighted problems due to the low repeatability of tests on human subjects. A possible solution is discussed in the second chapter.
Morover, the attention was focused on the data logger too. A new version of the Voicecare was developed: the microcontroller-based board used on the first version of the voicecare has been replaced with a more powerful one (Arduino Due); the new microcontroller allows to perform a real-time estimation of some vocal parameters extracted from the acquired signals.
The input section was improved with a programmable gain amplifier, useful to reduce the problems related to the saturation of the input channels. Moreover, a temperature and humidity sensor was implemented in order to obtain more information of the ambiental conditions during the monitoring session.
The improved version of the data logger has been tested in different conditions, in order to calibrate the acquisition algorithm and define its accuracy in the real time parameter estimation.

The second chapter deals with the development of a system that could provide a stable reference for testing and characterize vocal monitoring devices.
During the contact sensor tests discussed in the first part of the first chapter it emerged that these tests, performed on human subjects, do not allow to properly characterize the selected contact sensors: the high variability of the results did not provide enough information to properly discriminate a good contact sensor from a poor one. The main problem was the frequency response determination because it is not known how contact sensors interact with the human body, how they modify the mechanical properties of the tissues. It is important to understand these mechanisms because they introduce a filtering effects on the skin vibration.
The study of the voice frequency content and the related parameters is useful in the diagnosis of several vocal pathologies; for this reason it is important to asses how the sensor affects the phonatory system mechanical properties.
Based on these premises a system has been developed, a mechanical equipment that mimics the phonatory system, and what it generates: the acoustical voice emission and the vibration of the neck tissues. The developed system can act as a stable reference useful to perform repeatable measurements, in order to compare different contact sensors.
The system also embeds a sensor dedicated to determine the load effect of the transducer under investigation, a parameter that describes the impact of the contact sensor on the skin vibration.
The system has been tested, and its response has been compared with the response of the human body in order to evaluate its capability to mimic it, and the results are presented and discussed.

Some of the results presented in this dissertation has been published in [10-13].

Chapter 1

Voice monitoring acquisition system

1.1 Introduction

Voice disorders are increasingly affecting those professional categories that make extensive use of voice in a sustained way (teachers, singers, call-center operators, actors, salesman). This kind of voice use in working days may cause significant changes in vocal parameters with respect to non-working days [2], and the occurrence of voice disorders may cause mandatory absence from work in order to recover [3]. Various studies are still investigating the causes of these disorders, wich include different levels of vocal dysphonia, vocal fold nodules and related patologies [1], [4], [5].

The instrument used to monitor the vocal activity of voice professionals during a normal working day (long-term voice monitoring) is called "vocal dosimeter". It allows to quantify voice use as well as to relate voice parameters to voice disorders. Several studies have already dealt with the long-term voice monitoring [6], [7] and different vocal dosimeters have been developed and tested. Since the research in vocal disorders needs large data sets to test the parameter reliability, the interest in such kind of devices has been increasing [8]. The goal is to find the relationship between the way people uses their voice and the risks of disorders [9, 10]. Most of these devices record the vocal signal by means of contact sensors, that allow to minimize the effects of other sound sources in the environment besides the voice. By recording the vibration at the jugular notch (which is directly related to phonation)

the acquisition is focused on the signal of interest, reducing the influence of the external noise.
Contact microphones are able to record the voice in a clean way, even in disturbed environments like call centers and classrooms, and allow the estimation of signal amplitude-related parameters, such as the sound pressure level (SPL), and others related to the fundamental frequency such as the cepstral peak prominence (CPP).
The signal acquired at the output of the contact sensors is the vibration of the vocal folds, transmitted and filtered by the neck tissues. The source of this vibration is the same process that originates "the real voice" (the acoustic emission radiated from the mouth), and it has a lot of similarities with it but is not affected by the filtering effects of the vocal tract [19], which instead affects the acoustic emission of voice recorded with air microphones.

Several previously developed devices used an accelerometer as vibration transducer to detect the voice-related vibration signal at the jugular notch. For example, the National Center for Voice and Speech (NCVS) developed a research vocal dosimeter [11], [13] equipped with an accelerometer. Two commercial devices (no longer available) were also developed: the VoxLog, which was a commercial vocal dosimeter designed at the Linkopings University of Sweden [14, 15], and the Ambulatory Phonation Monitor (APM 3200), developed at the Massachusetts General Hospital [16-18]. Both of them were equipped with a miniature accelerometer too. Additional development has been carried on the APM 3200, but the device is no longer commercially available.
Unlike these devices, the Voicecare, recently developed at Politecnico di Torino, is equipped with an electret throat microphone (laryngophone) designed for helicopter pilots [19], [20]. This choice was made in order to use a low-cost transducer already suited for voice application. This chapter deals with the improvement of this device, the transducer and the data-log section.

As stated before, many research group have developed and used contact sensor-based vocal accumulators, especially in ambulatory monitoring [13, 19, 22, 23, 26, 28, 29], but there are no clear guidelines to follow in the selection of the right sensor. Guidelines for air microphone are instead well-established and used (see [21] for an overview).

The purposes of the first tests were:

- to evaluate the responses of different contact sensors, in order to select the best one to be used for voice monitoring;
- to provide a guideline for selecting contact sensors suitable for voice monitoring.

For these reasons, the performance of four contact sensors has been compared. For each sensor, three main characteristics were considered: the frequency response, the sensitivity to the background noise and to body movements. Specific tests procedures were established, starting from the tests used in [27] in order to obtain information on those characteristics. The the condition circuitry specifically designed for each sensor was taken into account too.

Alongside, the attention has been focused on the data logger and its improvement too. The existing Arduino program used on the original data logger was improved in order to obtain a real-time estimation of the fundamental frequency of the acquired vocalization, and a programmable gain amplifier was implemented on the input channels. Also, a compact air microphone (useful for the calibration and to record the acoustical ambient noise) and a temperature and humidity sensor were implemented. The system was electrically and acoustically tested.

1.2 Contact sensors performance comparison

The contact sensors selected to be tested and evaluated were two electret condenser microphones (different in size), an accelerometer and a piezoelectric transducer. From now on, when we talk about contact sensors, transducers or simply sensors, we are referring to the contact sensors under examination.
The idea was to use test subjects as reference: everyone would have to repeat the same tasks alternatively wearing the various transducers, in order to compare the responses of the transducers in the same conditions. The tasks were defined to imitate the activities normally performed during a vocal monitoring session, and will be explained in section 1.2.1. Every task has been repeated three times in order to

consider the repeatability.
The bigger electret condenser microphone (Midland MIAE38, ECMb) weighs about 4 g and has a diameter of 2.4 cm; the smaller one (Adafruit 1935, ECMs) has a nominal weight of about 0.17 g, and a diameter of 4.4 mm. The dimensionis of the accelerometer (Knowles BU-21771-000, ACC) is 7.8 x 5.5 x 4 mm, and its nominal weight is 0.28 g; the piezoelectric transducer (PIEZO), a contact microphone typically used on acoustic musical instruments, has a diameter of 3.2 cm and weighs about 3.6 g (see left-side of figure 1.1). The ECMs and the PIEZO are the cheapest (less than 5 Euros), the ECMb is a little less cheap (about 15 Euros) and the ACC is the most expensive (about 35 euros).
Two-channel data acquisition system has been arranged (see right-side of figure 1.1), capable of acquire two different voltage signal at the same time.
A normal air microphone (Behringer ECM8000) was connected to one channel and placed at 17 cm from the test subject's mouth, while one of the investigated sensor was connected the the other channel and attached to the test subject's jugular notch. The air microphone has been used as a reference and it was useful to control the acquisition and the performed tasks. It exhibits a flat response in the 60 Hz ÷ 2 kHz frequency range, and a maximum change of 1.5 dB in the 2 kHz ÷ 10 kHz frequency range. An external phantom power supply (Behringer PS400) was used to provide the polarization voltage (12 V) to the air microphone.

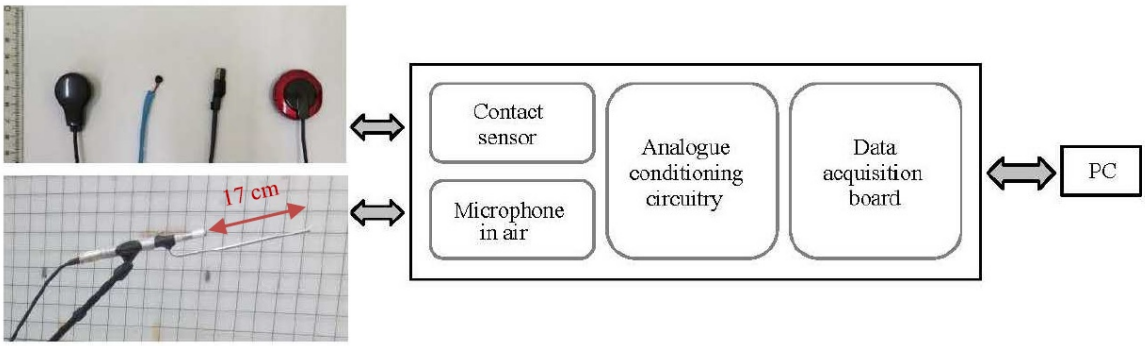


Fig. 1.1 Measurement chain used to test the contact sensors.

Five conditionig circtuits were designed, one for the air microphone and one for every contact sensor. Every circuit consists of a preamplifing section and a filtering section (a band-pass filter with 10 Hz ÷ 20 kHz bandwidth); these circuits also provide the voltage supply for the ACC and the phantom power for the ECMb and

the ECMs. The schemes of these circuits are reported in Appendix B.
the analog signals are then sampled by a data acquisition board (NI USB-6211) at 44kSa/s sampling rate with 16 bit resolution. The acquisition board is controlled by a LabView Virtual Instrument (VI) that runs on a personal computer, on which the data are stored. These data are then post-processed using MatLab scripts.

1.2.1 Testing procedures

The sound absorbing chamber of Politecnico di Torino was used to perform the tests; the measurement set-up is presented in figure 1.2. The two selected test subjects were a male (31 years old) and a female (25 years old), neither visual nor hearing impaired. The chamber background noise level was measured with a calibrated class-1 sound level meter (NTi Audio XL2), and it was found to be 26.2 dB.
The sensors have been attached at the test subjects' jugular notch by means of a surgical band. The attachment point on the test subjects' neck was marked, in order to place every sensor in the same position. For each sensor, subjects were asked to perform four different tasks; each task is a test useful to define characteristics of the sensors.

- test n.1: sustain the vowel /a/ for 10 seconds in front of the air microphone, at four different levels, from a very low intensity (like a whisper) to a very high intensity (almost like a scream). The vowel /a/ was chosen among other vowels because it exhibits the best repeatability in this kind of tests [24]; this task reproduces the calibration task which is normally performed to calibrate the contact sensor channel on the Voicecare.

- test n.2: keeping the mouth in front of the reference microphone while rotating the body from left to right; stay in silence for 10 seconds and then repeat the italian passage "ninna nanna nonna Anna" three times at comfortable voice level. That sentence was chosen because it is an all-voiced passage, useful to clearly distinguish between the non-voiced and the voiced parts of the acquired signal. This task was meant to understand how the body movement affects the signal acquisition with the different contact sensors.

- test n.3: stand in front of a sound source (NTi Audio TalkBox) set to generate a 60 dB white noise at a distance of 1 m. The position assumed by the test subjects, with the chin leaned on the sound source, allows to radiate the sound directly on the contact sensor attached to the neck; the sound pressure level measured very close to the contact sensor is 88 dB. For the first 5 seconds of the test, the sound source is off and the test subject has to remain silent. The acoustic source is then turn on, and the test subject remains silent for another 5 seconds; for the last 10 seconds the test subject repeats three times the passage "ninna nanna nonna Anna" at increasing voice levels. Then, the procedure is repeated with the sound source set to generate a 70 dB white noise at 1 m, that corresponds to 98 dB near the contact sensor. The goal of this test was to quantify the sensitivity of the various contact sensors to sound waves. Voice monitoring sessions often occur in places filled with background noise, like classroms ad call centers; in these situations, it is important that the contact sensor is not influenced by the acoustical background noise.

- test n.4: sustain the /a/ for 10 seconds, trying to produce it at the same tone for every acquisition.

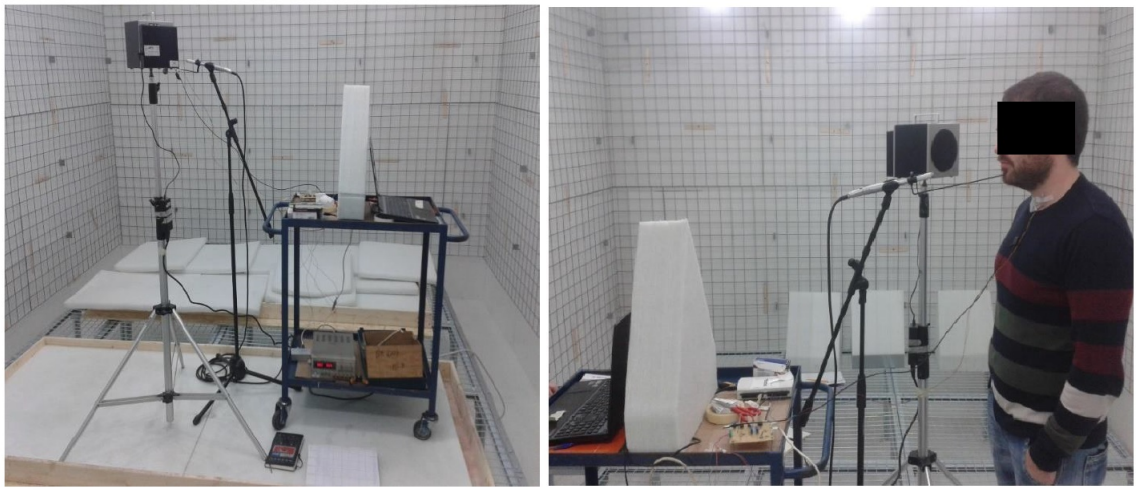


Fig. 1.2 Measurement setup: air and contact microphone, acoustic source, conditioning circuits, acquisition board and PC.

1.2.2 Calibration procedure

Measuring the SPL of the vocalization by sensing the skin vibration (induced at the jugular notch by the vocal folds movement) requires a calibration procedure, in order to relate the acquired vibration to the radiated sound. Generally speaking, every device which performs a measurement of a physical quantity needs a proper calibration.
The acquisition system used for these tests has two channels, one for the air microphone and one for the contact sensor, like the Voicecare. The calibration procedure of its contact sensor channel consists in repeating the /a/ vowel in front of the air microphone at different intensity, wearing the contact sensor, in order to relate different sound pressure voice levels to the intensity of the vibration at the base of the neck [24]. This procedure is similar to those used with other vocal dosimeters [12, 13].
Firstly, the calibration constant K_{mic} of the air microphone channel has been estimated by coupling the air microphone to a sound pressure calibrator B&K 4230, which provides a nominal pressure of 1 Pa @ 1 kHz. The calibration constant is

$$K_{mic} = \frac{SPL_{ref}}{Vrms_{mic}} \tag{1.1}$$

where SPL_{ref} is the level emitted from the calibrator and $Vrms_{mic}$ is the root mean square of the signal amplitude at the output of the microphone. It is calculated on 30 ms long frames of the signal. The same formula allows to obtain the SPL_{mic} from the air microphone, which is necessary to perform the contact channel calibration:

$$SPL_{mic} = K_{mic} * Vrms_{mic} \tag{1.2}$$

Then, the function that relates the contact sensor signal to the voice sound intensity can be obtained by using the calibration procedure described above: the subject under monitoring repeats the /a/ vowel at a fixed distance from the air microphone (17 cm), with the contact microphone attached at the base of the neck. The vocalization acquired with the two transducers is divided in 30 ms frames, and the $Vrms_{mic}$ and the $Vrms_{cont}$ (the root mean square of the voltage output of the contact sensor) of every frames are obtained. The $Vrms_{mic}$ is used to obtain SPL_{mic}, and then the SPL_{mic} - $Vrms_{cont}$ function is obtained by relating the two values from the same frames. The identified functions (top charts in figure 1.3) allow to express

the results of the tests (the vibration acquired at the base of the neck) in terms of Sound Pressure Level SPL (dB), referred to the distance of 1 meter. It is a logarithmic function which needs the computation of two parameters, K_0 and K_1:

$$SPL_{mic} = K_0 + K_1 * \log_{10}(Vrms_{cont}) \tag{1.3}$$

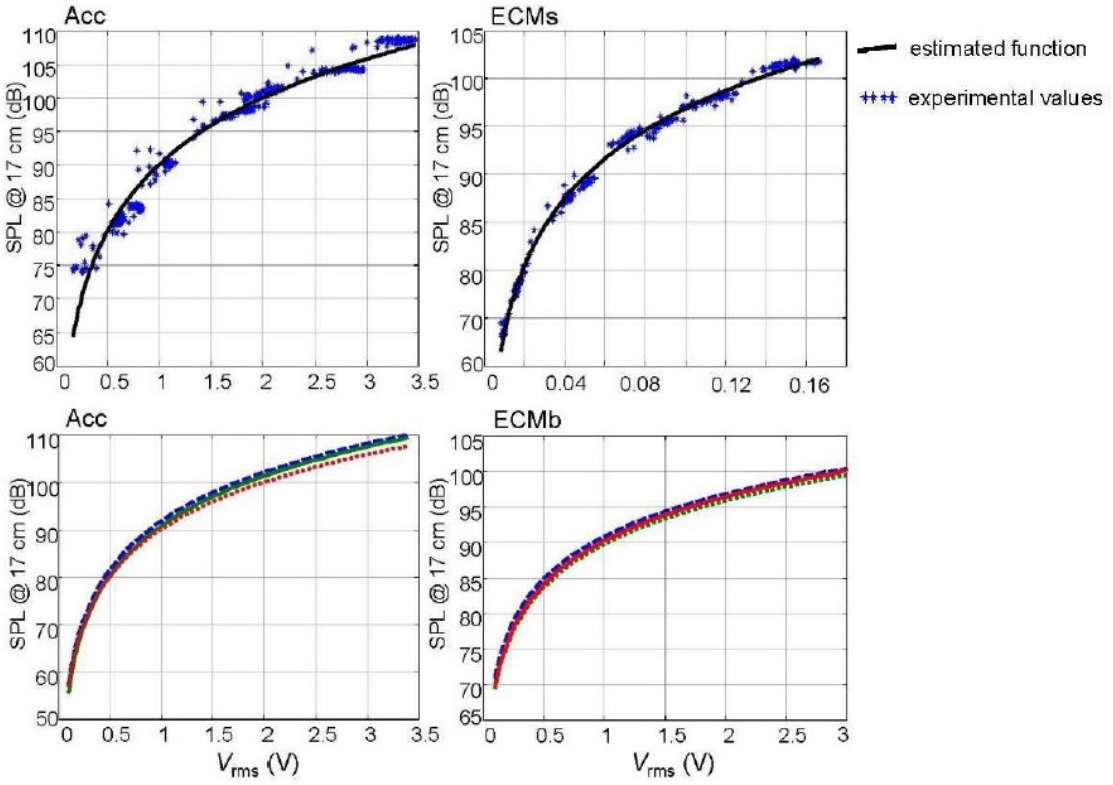


Fig. 1.3 Calibration functions estimated from the accelerometer (Acc) and the smaller ECM (ECMs) (top charts); repeatability of the calibration procedure with the vowel /a/ for the accelerometer and the bigger ECM (ECMb)(bottom charts).

This procedure is valid only for the subject who performs it, and just for the configuration used for it, because it depends on type, shape and dimensions of the contact sensor, its position at the base of the neck, and on the subject's physical characteristics.

For every sensor, the procedure has been repeated in order to obtain different calibration functions. The comparison of them does not highlight significant differences (bottom charts in figure 1.3), confirming the repeatability of the calibration proce-

dure stated in [24]. However, this fact does not help in the sensors performance comparison because similar results were obtained for all sensors.

1.2.3 Sensitivity to the background noise

In order to quantify the sensitivity of each contact sensor to the acoustical background noise, test n. 3 was performed; the obtained results are summarized in figure 1.4. As highlighted in [25], it is important to perform this test with the contact sensor attached to the neck of a test subject: the skin and the other neck tissues change the mechanical response of the sensor (which would be different if it was attached to a rigid surface), so it is important to evaluate the acoustical background noise sensitivity of sensors in that situation, when they are attached to the human body.

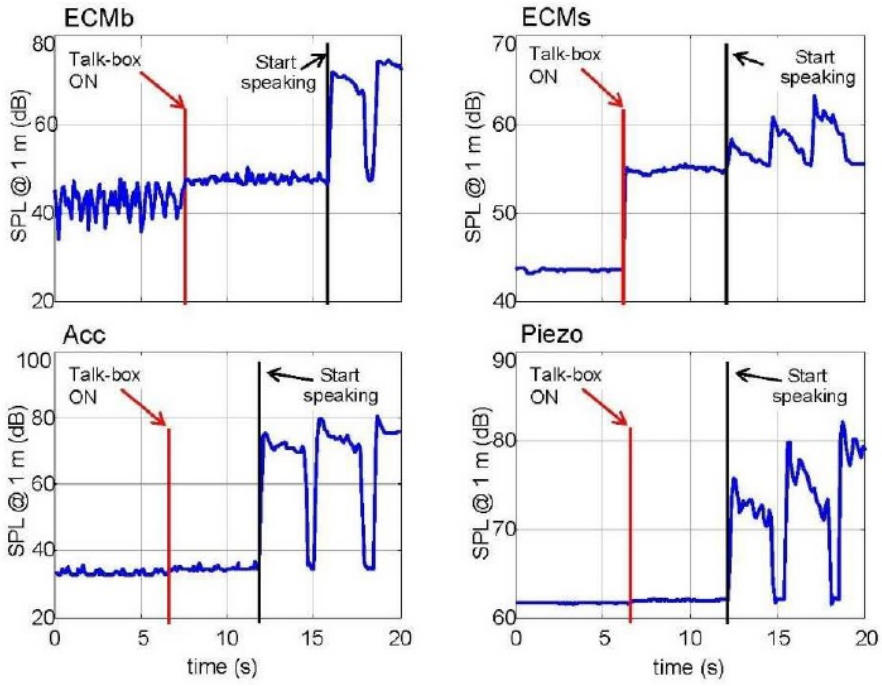


Fig. 1.4 Air-borne effects on the sound pressure level (SPL) referred to the distance of 1 meter, for each contact sensor. Noise intensity in correspondence to the sensor worn by the subject was 98 dB.

In order to quantify that particular sensitivity, for every sensor two parameters have been considered: the internal noise floor, and the signal level obtained in the

first part of the test n. 3. The gap between them is a parameter which expresses the sensitivity to the acoustical background noise. The smaller ECM (ECMs) is the most sensitive, with a gap of about 12 dB; the bigger ECM (ECMb) exhibits a reasonable difference of 5 dB. The accelerometer and the piezoelectric transducer are basically insensitive to the sound box emission, with a gap smaller than 2 dB, confirming and extending the results found in previous investigations [19, 25]. This test has also provided information on the internal noise floor: it is almost 40 dB for all the sensors except for the piezoelectric transducer; its noise floor is higher than 60 dB. The accelerometer has the lowest internal noise floor, a little less than 40 dB.

1.2.4 Tissue-born effects

As already stated, the contact sensor is a non-intrusive way to record the voice: it can be worn without interfering with the normal daily activities of the subject under monitoring. This is one of the reasons why contact sensors are used in the long-term voice monitoring. But normal daily activities may involve movements (think about a professor who teaches a class). The test n. 2 was designed for this purpose: investigate the tissue-borne effects due to body movements.
The results show that the signal of all the contact sensors contains artefacts due to body movements. These artifacts disturb the voice acquisition, contaminating the data; this fact was already highlighted in [12] for the accelerometer. This test shows that the artifacts have peak values similar to those obtained during the normal speaking activity.
The frequency analysis of the acquired signal has shown detached frequency components from body movements (below 30 Hz) and speaking activity. As suggested in [24], the acquired samples have been post-processed through a high-pass filter, with the cut-off frequency set to 50 Hz.
Applying this post-processing procedure, the artifacts introduced by body movements are made negligible. The amplitude gap between filtered and unfiltered signal is shown in the top chart of the figure.
Differences in SPL amplitude are more noticeable in the bottom chart of the same figure.

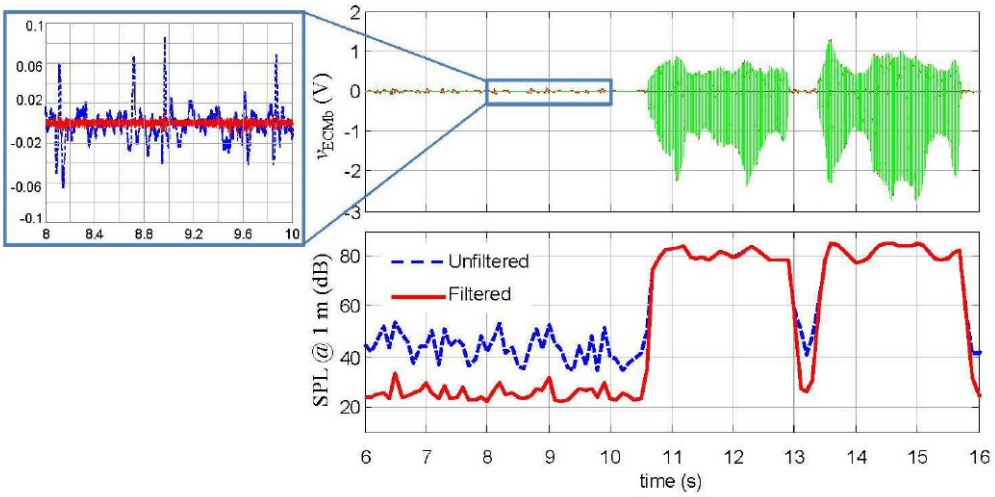


Fig. 1.5 Voltage signal at the output of the ECMb (upper chart); highpass filter effect on the sound pressure level (SPL) estimated from the signal (lower chart).

1.2.5 Frequency response

The signals acquired during test n. 4 have been processed according to a Fast Fourier Transform (FFT) algorithm, in order to estimate the frequency components of the vibration signal acquired with the various sensors. Frame lengths of 200 ms were processed through the FFT algorithm, thus obtaining a frequency resolution of 5 Hz; a Hamming window was applied to each processed frame in order to minimize the effects on the non-coherent sampling.

Two examples of the obtained spectra are shown in figure 1.6, where the grey lines refer to the air microphone signal spectra, and the darker lines are contact sensor signal spectra. The presented graphs were obtained with the accelerometer (left-side chart in figure 1.6) and the ECMb (right-side chart in figure 1.6). The reported values were normalized with respect to the amplitude of the highest frequency component.

The spectral development of the /a/ signal acquired by the air microphone (grey lines) follows the typical vowel spectra, which can be defined as the spectra of the glottal source filtered by the vocal tract. The source filter theory assumes that the glottal spectra exhibits a fundamental frequency and a series of harmonic components at decreasing amplitude.

The voice signal exhibits a continuous spectra with multiple peaks, which are the so

called formant frequencies. The vocal tract and the oral cavity act like a resonator on the pressure signal originated by the glottis, creating the peaks on the spectral envelope. Some of that are modulated in the speech process, which are the frequencies that distinguish one vowel from another.
The signals acquired by means of contact sensors are not subjected to the filtering effect of the vocal tract, hence the spectra show a decreasing trend similar to the pressure signal originated by the glottis.

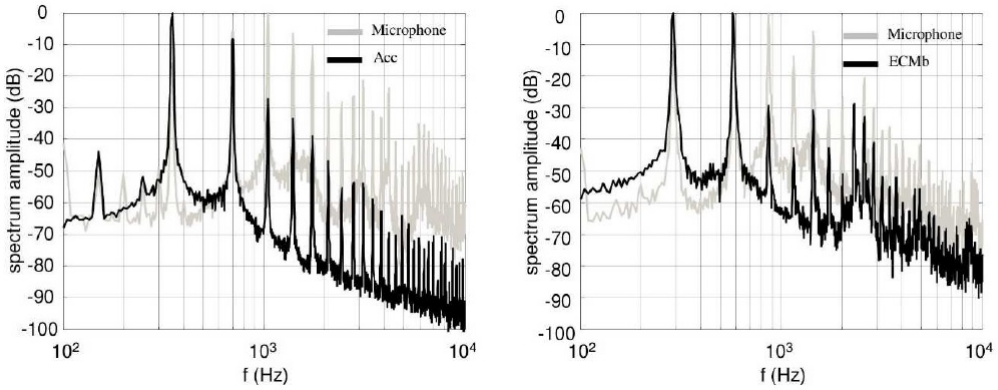


Fig. 1.6 From left to right: spectra of the signal at the output of the microphone and the accelerometer (Acc) channels; spectra of the signal at the output of the microphone and the bigger ECM (ECMb) channels.

It is important to notice that even if the test was repeatedly performed by the same subject, trying to emit the /a/ vowel at the same tone for every tested sensor, the obtained spectra (even the air microphone ones) were always different, so it was difficult to make a quantitative comparison of the results obtained with different sensors. Furthermore, the differences in the contact sensor spectra reported in figure 1.6 can not be explained only by different frequency responses, because the two air microphone spectra are different too. This has to be related to the way that every sensor perturbs the system under measurement, in this case the vocal tract.

By the way, the amplitude spectra obtained for the signal recorded by the two channels have been compared, in order to point out at least some quantitative considerations. The two charts in figure 1.6 show that the air micropohone signal spectra is characterized by components higher than -60 dB for frequencies up to about 10 kHz. By using this amplitude value (-60 dB) as a threshold for the contact sensors signal

spectra too, the accelerometer exhibits a frequency range slightly higher than 3 kHz, while ECMb shows a frequency range of about 4 kHz. A similar result was obtained for the piezoelectric transducer. It was not possible to obtain a meaningful characterization for the ECMs because of its high sensitivity to the voice acoustical emission, that corrupted the vibration acquisition.
These values does not refer to the "real" frequency range of the sensors, but express the frequency range of the neck vibration (related to the phonation) that every sensor is able to acquire. Being the real bandwidth of all the investigated contact sensors of about 10 kHz, a low-pass filter effect is noticeable, which has to be ascribed to the physiological elements of the channel.

1.2.6 Conclusions

These tests did not provide enough useful information to properly discriminate and rate the contact sensor to be used for vocal monitoring. The calibration test (test n. 1) point out that the calibration procedure works well for different sensors, but this did not help to evaluate the quality of one sensor towards another.
The test n. 2 point out that the sensitivity to the body movement (tissue-born effect) can be compensated by using a digital filter, and therefore does not depends on the contact sensor.
The test n. 3 was the only one which provide some good results in evaluating the property of the contact sensors. The fact that the contact sensor is worn by the subject during the test is important because allows to keep into account the effects of body tissue attenuation and noise introduced by other body-born events, like breathing and heartbeat. The accelerometer and the ECMb (already used on the Voicecare) have been found the bests in terms of internal noise floor and the sensitivity to the background noise, when worn by the subject under monitoring. Actually, the accelerometer exhibits a better behaviour in this test, but it is more expensive than the ECMb and its implementation requires more work.
The data collected in the test n. 4 has showed that no affordable comparison can be performed on a human subject for two main reasons: the low repeatability, and the impossibility to know the response of the system under measurement, unloaded.
Every time the sensor has been changed and replaced with the next one to test, and every time the test subject tried to repeat the same vocal task, the results were

different. It seems to be nearly impossible to repeat a vocalization at the exact same frequency and intensity, for a subject who is not trained. And for this reason it was impossible to correctly compare the response of the different contact microphone.
As already stated before, the contact sensor attached at the base of the neck seems to modify the measured system and change the frequency content, in the vibration at the jugular notch but also in the acoustical emission radiated from the mouth. This would not be a problem if the original response of the "unloaded" jugular notch tissues was known, but it is not possible to perform such a measurement: it would entail the use of a contactless vibration recording apparatus, like a laser vibrometer, but it would be a very complex process and it would not give sure results. But still, it will be very useful to be able to measure and rate the invasiveness of a contact sensor, in terms of how much it modifies geometry and mechanical properties of the neck tissues.

Apparently, it seems that the human body is not a stable test system; a different way to test contact sensor for vocal monitoring will be discussed in chapter 2.

1.3 Data logger improvement

Alongside the tests to evaluate and select contact sensors, the improvement of the acquisition system used for vocal monitoring was also been carried on. The basis on which this work started was the Voicecare, already mentioned in the Preface. It consists in an arduino uno-based portable device equipped with two acquisition channels, one for the contact sensor and another one for an air microphone, whose only purpose is the calibration of the contact sensor channel. The Voicecare acquires only raw vocal data, which means that the data logger simply samples and stores the signal produced by the transducers. The data are then processed (using ad-hoc Matlab scripts) in order to obtain the contact sensor calibration function first, and then the SPL and fundamental frequency of the vocal data, and possibly other vocal parameters useful to diagnose voice disorders.
The vocal monitoring performed with the Voicecare consists in two different procedures: calibration and acquisition. The two procedures are implemented through two different operating modes.

The calibration mode uses two channels, one for an air microphone and one for a contact sensor (the two transducers are the Behringer microphone and the bigger ECM contact sensor (ECMb) used for the tests described in section 1.2). The air microphone acquires the phonation radiated from the mouth, and the contact sensor acquires the vibration (skin acceleration level, SAL) at the jugular notch. The two analog signals are then sampled by the ADC and stored on an SD card embedded in the arduino board.
The microphone signal allows to obtain sound pressure levels of every frame recorded in the calibration mode. A calibration function is then estimated by postprocessing the data: the samples are divided in frames, and then the procedure described in section 1.2.2. is used.
The calibration function allows to obtain the voice SPL of the subject under monitoring only from the signal of the contact sensor attached to his/her jugular notch (SPL_{cont}).

The acquisition mode uses just one channel, the contact sensor one, to acquire the vibration signal at the jugular notch. The analog transducer signal passes through some conditioning circuits before being sampled by an ADC and stored by the arduino board on a microSD; the gain on the two channel is fixed, and this fact introduces some saturation issue, when the signal is too strong for the arduino analog input.
The signal is then post-processed: it is divided in 30 ms frames and the SPL_{cont} for every frame is obtained:

$$SPL_{cont} = K_0 + K_1 * \log_{10}(Vrms_{cont}) \tag{1.4}$$

where K_0 and K_1 are the parameters obtained from the calibration procedure postprocessing.

In order to improve the device, some modifications have been done. Firstly, the Arduino Uno was replaced by an Arduino Due: the microcontroller on the new board has a double buffer, so it is possible to use one buffer for the data acquisition and the other one to calculate parameters on board, in real time, from the acquired

signal. The greater processing power has been used to process the data in real time, to divide the signal in frames and obtain the V_{RMS} value of the contact sensors output in every frame (this operation for the first version of the Voicecare is performed during the postprocessing); this was the first step to eventually obtain the SPL from the same signal in real time. The fundamental frequency f_0 can be obtained in real time too by the direct analysis of the recorded signal. For the first version of the Voicecare, the fundamental frequency is estimated in the postprocess, by means of an autocorrelation algorithm.

These implementations are important because the subsequent step of the device improvement should have been the installation of a small screen, so the user could have check his vocal parameters in real time, in order to control and possibly correct his/her own vocal behaviour.
Moreover, the new board allows to sample the signal at a higher frequency due to the greater ADC resolution (12 bits for arduino due instead of 8 bits for arduino uno). This could allow the data logger to be used for contact microphone testing to, which require a high sampling frequency to obtain a high level of detail on the signal acquired with the tested sensor.
The correct operation of the device in computing Vrms and fundamental frequency in real time was electronically tested. This implementation will be discussed more in depth in section 1.3.1.

A Programmable Gain Amplifier (PGA) has been implemented on the input channels. It is a voltage amplifier which can be controlled by the arduino microcontroller, and it could resolve the saturation issue that usually affects the acquired signals by automatically scaling the gain in respect of the magnitude of the transducers output. The first Voicecare version is equipped with a simple voltage amplifier with fixed gain on every acquisition channel; this helps in detecting soft voices (which imply low vibration levels), but reduces the useful acquisition range because a saturated frame is useless for the analysis. The PGA, if required, can low the gain to 0 if the input signal is strong, and raise it if the signal is low.
The PGA has two channels, and the gain selection is independent. This means that one channel can be used for the contact microphone, and the other one for the air microphone. Its implementation will also be discussed more in depth.

Other improvements, not included in this dissertation, were the implementation of another air microphone to sense the acoustic noise floor in the room when the monitoring takes places, and the implementation of a temperature/humidity sensor.

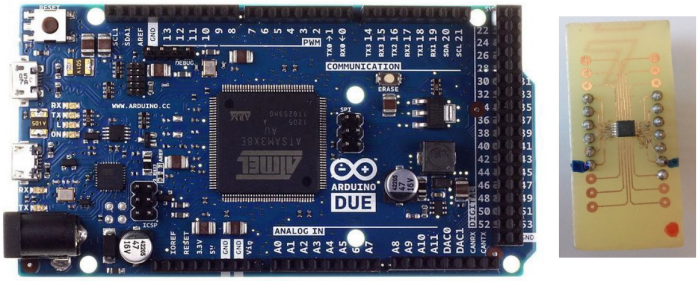


Fig. 1.7 The Arduino Due board used for this project (left), the PGA (right).

1.3.1 Gain automation ad real-time parameter estimation: electrical tests

The human speech can reach very different levels, from a whisper to a yell. In a microcontroller-based data logger like the Voicecare, the range of acquirable values is limited and the acquisition often presents saturation issues when the input is bigger than the acquisition range. In this sense, the gain automatization could be very useful to acquire weak and strong signals without saturation problems, by changing the gain level in respect of the input voltage.

The used amplifiers is the PGA112, which can be digitally controlled via SPI by the Arduino Due board. This model has 8 (0-7) binary gain levels: the gain for the $n-th$ level corresponds to the 2^n gain. The voltage Reference level is chosen by the user, we chose 1.65 V, the half of the scale of the arduino analog input. This input do not accept negative voltage values, so in this way it can sense even from a bipolar transducer.

The logic developed to control the PGA for this application computes the Vpp (peak-to-peak voltage) for every acquired frame, and then calculates the difference

among this value and the maximum voltage admitted by the arduino input channel and:

- if the difference is less than 40% of the reference level, the gain is shifted up;
- if the difference is more than 90% of the reference level, the gain is shifted down;
- if the difference is between the 40% and the 90% of the reference level, the gain is not shifted.

The information on the gain level of both the PGA channels is stored in the buffer dedicated to the parameter evaluation for the duration of a frame, and it is also written on the SD card with the rest of the acquired and estimated data. The Arduino Due is programmed to calculate the $Vrms$ of every acquired frame in real time: the $Vrms$ of the "raw" acquired signal is calculated, and then it is multiplied by the inverse of the gain level selected for that particular frame; the result is stored in the buffer for the duration of a frame and then in the SD card. In this way it is possible to obtain the real amplitude of the signal produced by the transducers, and to compensate the changes of the gain level in real time, during the calibration and the monitoring.
Therefore, the data about the signal and the gain level of every frame is available for an eventual postprocess.

Electrical tests have been performed, to validate the efficiency of the gain automatization and to test the real-time parameter estimation. The two channels of the data logger are identical, so the same tests were performed on both of them. The input of the tested channel has been connected to the output of a signal generator, which was generating a digital signal that reproduces the /a/ vowel.
The data logger acquired the signal and estimated the $Vrms$ and the $F0$. Those values have been then compared to the ones read on the generator (the signal was periodic and stable). The test was repeated setting different $Vrms$ and $F0$ on the signal generator, in order to test the PGA operation too: it was important to verify the original amplitude computation for every gain level, and at different frequencies. Since the screen on which the parameters should be visualized has not be implemented yet, the data and the parameters are estimated in real time, stored on the SD

Table 1.1 REAL-TIME Vrms ESTIMATION VERIFICATION : *frequency* is the frequency of the test signal; *Vrms in* is the amplitude of the test signal; *Vrms out* is the amplitude calculated by the data logger; *gain* is the PGA gain level; $\Delta Vrms$ is the difference between Vrms in and Vrms out.

frequency (Hz)	Vrms in (mV)	Vrms out (mV)	gain	ΔVrms (mV)	ΔVrms %
100	19.76	0.00	7	19.76	-
100	23.67	0.00	6	23.67	-
100	25.62	8.00	5	17.62	68.77
100	35.46	32.40	4	3.06	8.62
100	62.99	62.80	3	0.19	0.31
100	100.33	99.80	2	0.53	0.53
100	178.78	178.70	1	0.08	0.04
100	385.95	384.20	0	1.75	0.45
150	19.73	2.30	7	17.43	88.34
150	23.68	8.90	6	14.78	62.42
150	25.60	15.40	5	10.20	39.84
150	35.46	32.00	4	3.46	9.74
150	62.98	62.90	3	0.08	0.13
150	102.27	102.00	2	0.27	0.27
150	178.72	178.40	1	0.32	0.18
150	374.06	372.00	0	2.06	0.55
200	19.74	2.30	7	17.44	88.35
200	25.60	9.00	6	16.60	64.84
200	25.60	13.80	5	11.80	46.10
200	31.52	29.10	4	2.42	7.68
200	62.96	62.90	3	0.06	0.10
200	102.25	102.00	2	0.25	0.25
200	182.62	182.20	1	0.42	0.23
200	377.92	377.00	0	0.92	0.24

frequency (Hz)	Vrms in (mV)	Vrms out (mV)	gain	ΔVrms (mV)	ΔVrms %
250	19.74	2.20	7	17.54	88.86
250	24.43	8.00	6	16.43	67.26
250	25.60	15.20	5	10.40	40.62
250	43.32	38.40	4	4.92	11.36
250	62.96	63.00	3	-0.04	-0.06
250	102.24	102.00	2	0.24	0.24
250	188.60	182.00	1	6.60	3.50
250	377.84	377.20	0	0.64	0.17
300	19.74	2.20	7	17.54	88.86
300	25.60	15.20	5	10.40	40.62
300	35.44	32.00	4	3.44	9.69
300	62.95	63.00	3	-0.05	-0.07
300	102.24	102.00	2	0.24	0.23
300	180.64	180.30	1	0.34	0.19
300	375.93	375.20	0	0.73	0.19
350	24.64	7.10	6	17.54	71.19
350	25.60	15.20	5	10.40	40.63
350	43.32	38.30	4	5.02	11.59
350	62.97	62.90	3	0.07	0.11
350	102.26	102.30	2	-0.04	-0.04
350	180.67	180.40	1	0.27	0.15
350	376.00	375.20	0	0.80	0.21
400	23.66	5.40	6	18.26	77.17
400	25.60	15.00	5	10.60	41.40
400	35.43	32.00	4	3.43	9.68
400	62.95	62.90	3	0.05	0.08
400	102.24	102.00	2	0.23	0.23
400	180.63	180.30	1	0.33	0.18
400	375.91	375.20	0	0.71	0.19

card and then analyzed after the tests. The results are reported in table 1.1.

$\Delta Vrms$ values (the difference between the source amplitude and the calculated amplitude of the input signal) point out a good estimation of the source signal amplitude, when the input requires a level 4 gain or lower for 100 and 200 Hz signal, and a level 5 gain or lower for 250, 300, 350 and 400 Hz signal. This was not because of the PGA amplification, but it is due to the fact that when a high gain level is required, it is because of a very weak signal (mostly noise) which is not well-read by the ADC of the data logger. The usual signal that comes from a contact sensor or an air microphone is usually stronger, and moreover a pass-band filter

Table 1.2 REAL-TIME FREQUENCY ESTIMATION VERIFICATION: *mean f* is the mean frequency estimated by the data logger on a 15 seconds acquisition; *ref f* is the frequency of the signal read on the signal generator; *max* Δf is the maximum difference between *ref f* and *mean f* on a 15 seconds acquisition; *mean* Δf is the average difference between *ref f* and *mean f* on a 15 seconds acquisition; *standard error* is the standard deviation of the frequency values estimated by the data logger.

mean f (Hz)	ref f (Hz)	max Δf (Hz)	max Δf percentage	mean Δf (Hz)	mean Δf percentage	standard error (Hz)
100.00	100	0.00	0.00%	0.00	0.00%	0.00
150.37	150	1.52	1.01%	1.13	0.75%	1.19
200.00	200	0.00	0.00%	0.00	0.00%	0.00
250.00	250	0.00	0.00%	0.00	0.00%	0.00
303.03	300	3.03	1.01%	3.03	1.01%	3.04
344.83	350	5.17	1.48%	5.17	1.48%	5.19
400.00	400	0.00	0.00%	0.00	0.00%	0.00

could be implemented on the input channels of the data logger, in order to increase the signal-to-noise ratio.
More in general, a high amplification it is not recommended due to the distortion and noise amplification that is added to the original signal.

In order to verify the real-time frequency estimation, the data logger was tested with the same signal used for the previous test, at different frequencies but with a fixed 0,1 mV amplitude, and with a 15 seconds duration. Then, the values estimated by the data logger were compared to the reference ones, read on the frequency generator.
The fundamental frequency $F0$ is estimated on board by the data logger by means of an algorithm derived from the postprocessing algorithm used for the first version of the Voicecare. The original alghorithm (discussed and tested in [19]) uses a time-domain autocorrelation processing, and it was ported to arduino language in order to perform the real-time analysis. Results are reported in table 1.2.

The highest *standard error* is nearly 5 Hz, which is acceptable because it is almost the same value obtained for the postprocessed data for the validation of calibration procedures and uncertainty estimation performed on the prototype version of the data logger [24].
The sensitivity of the data logger in estimating the fundamental frequency has been tested too, by setting the signal generator to generate a frequency sweep of the

same signal used for the other tests. The sensitivity in the fundamental frequency evaluation was found to be 4 Hz.

1.3.2 Saturation issue troubleshooting

After the electrical test, the improved device has been tested in a situation more similar to a classical voice monitoring session: a test subject has performed a vocal task in front of the air microphone with the contact sensor attached to the jugular notch, with the system in the calibration mode. The vocal task consists in producing the /a/ vowel at different intensities and reading a passage at a comfortable level. This was intended to test the PGA efficiency in a real voice data acquisition. The acquired signals are reported in figure 1.8.

The 0 and the upper horizontal lines in the graphs represent the ADC limits: when the signal reaches those levels, the input is saturated and the frame is not good for the analysis. The percentage of saturated frames (PS) is greater for the air microphone channel and it is very high, due to the fact that the speech have a very variable level even in a single word, due to the various unvoiced sounds produced (like the consonants). Moreover every little pause, even the intersyllabic ones, lowers the gain to 0 so the following frames are saturated.

This problem was resolved by implementing a parameter (INCMAX) which is the number of samples that the system waits before the gain is shifted up or down. The test has been repeated for different values of INCMAX, the results are reported in figures 1.9 and 1.10

Considering the results for the /a/ vowel and for the passage reading too, the incmax value that minimize the saturate frame percentage in both situation is incmax=4. The /a/ is important because it is used in the ad-personam calibration procedure, and the passage reading is similar to a real monitoring session, when the subject under monitoring emits voiced and unvoiced sounds.

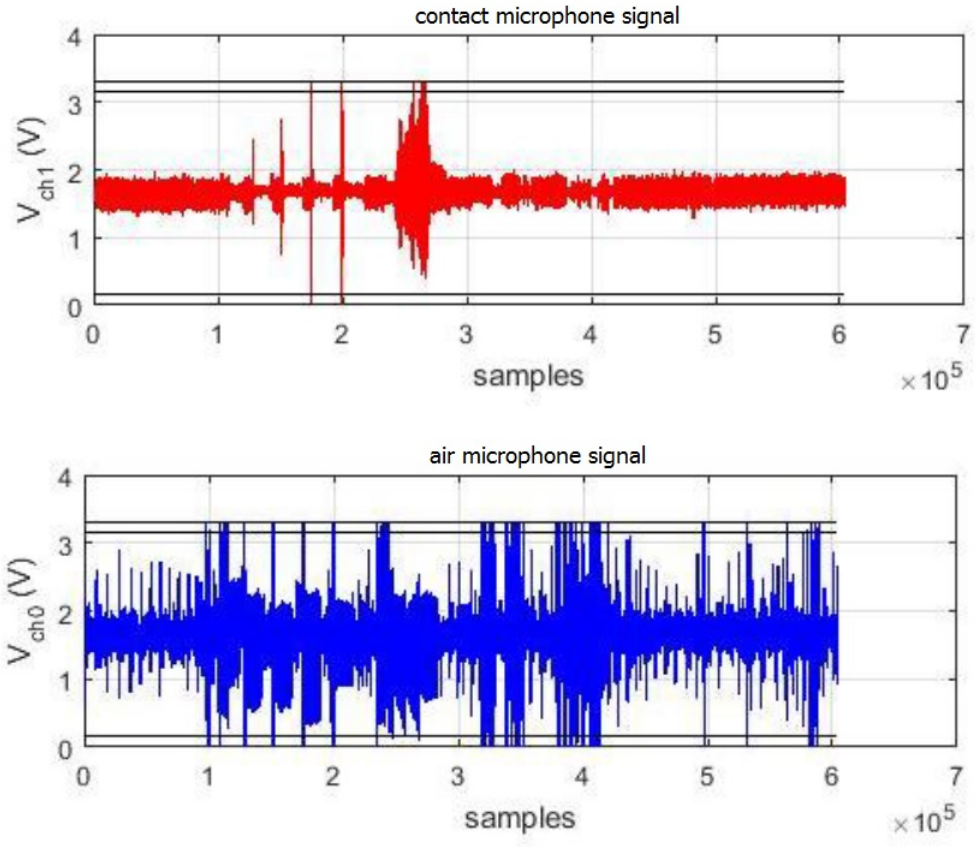
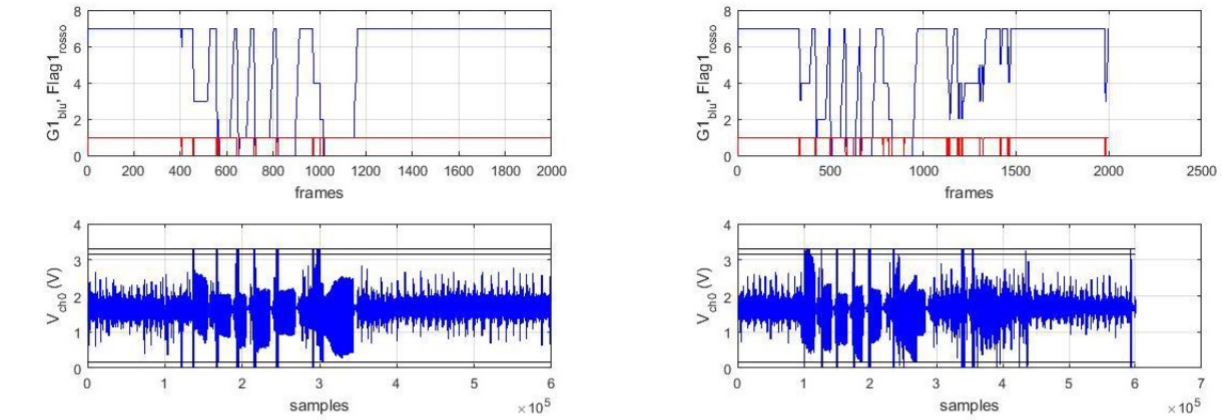


Fig. 1.8 Signals acquired with the data logger equipped with the PGA.

incmax=3 PS= 2,2545%: incmax=4 PS= 4,6477%:



incmax=5 PS= 4,3697%: incmax=6 PS= 2,2795

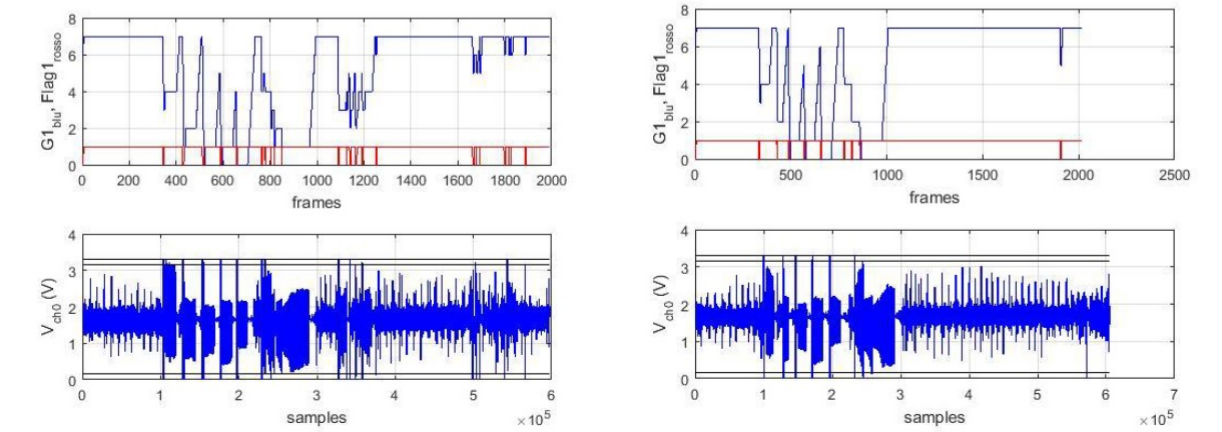
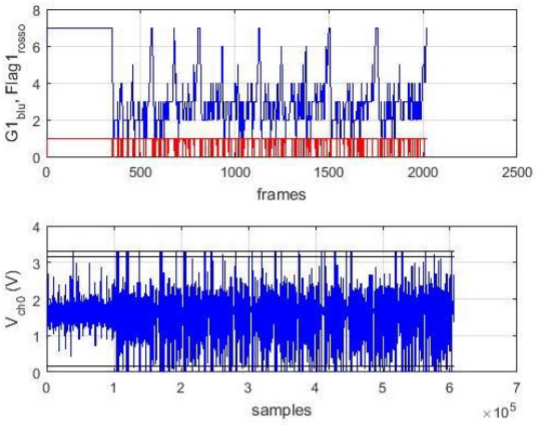
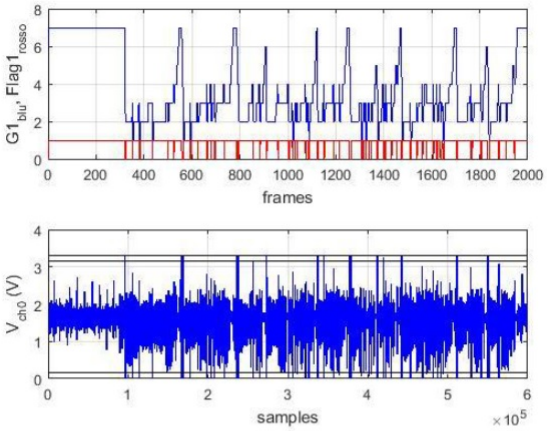


Fig. 1.9 Percentage of saturated frames for different values of the INCMAX parameter, for the /a/ vowel.

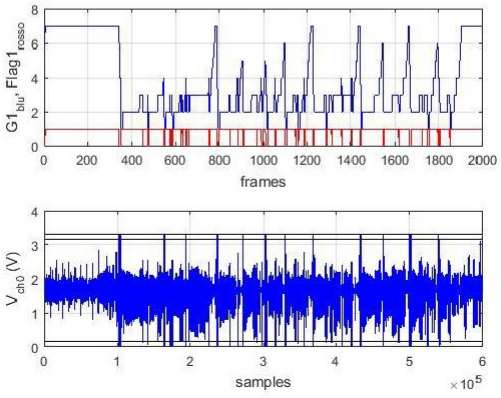
incmax=3 PS= 14,4483%:



incmax=4 PS= 4,4084%:



incmax=5 PS= 7,4612%:



incmax=6 PS= 6,6165%:

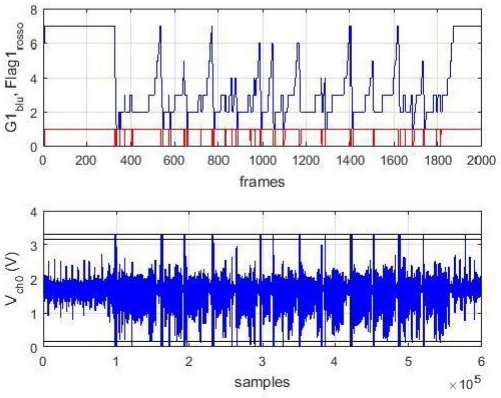


Fig. 1.10 Percentage of saturated frames for different values of the INCMAX parameter, for a passage reading.

1.3.3 Conclusions

All tests performed on the improved version of the Voicecare have produced good results: the real-time estimation of the $Vrms$ and $F0$ for the signal, divided in frames, works well, and the difference between the estimations and the real values are acceptable. The $Vrms$ estimation algorithm showed issues when the gain level was lower than 5, but such weak signal in a vocal monitoring session means no vocal activity, and a signal frame with no vocal activity contains no useful data and thus it is discarded.
The fundamental frequency estimation is acceptable for the purpose, in terms of standard error and sensitivity. The PGA operation effectively increase the data logger measurement range, and the implementation of the incmax parameter could really helps in overcoming the saturation issues occurred in the monitoring session performed with the first version of the Voicecare.

The increased computing power allows to control other sensor and interface, like the temperature and humidity sensor and a screen. The ambient data can be used to relate incorrect vocal behaviour to the conditions present in the room when the monitoring session took place. The possibility to visualize vocal parameters in real time could help to control and possibly to correct the vocal behaviour during the monitoring session. This could allow to use the vocal dosimeter not only as a research and diagnostic system, but also as a care instrument.

The algorithms for the vocal parameter evaluation are being also used on an experimental vocal monitoring application, designed to be used on a normal smartphone. The use of a smartphone seems to be easier than the development of a dedicated data logger, but the smartphone electronic is less manageable (gain settings, automatic digital noise reduction and compensation, filtering effects) and, more important, it introduces issues to ensure the metrological traceability: the air microphone used for the contact channel calibration on the Voicecare can be calibrated by means of a sound meter calibrator (a sound pressure traceable reference), but a smartphone integrated microphone can't be properly calibrated.
On the other hand, smartphones allow to collect a large amount of data because of their big circulation: the test subject could use his/her smartphone by simply downloading the app and using a contact sensor like the ECMb already used on the

Voicecare, which is already suited to be used for vocal application (it is a throat microphone designed for chopper pilots).

Chapter 2

Phonatory System Simulator

2.1 Introduction

As seen in the previous chapter, tests on human subjects are the usual way to study the response and the operation of vocal dosimeters, and to evaluate the uncertainty related to that type of measurement [7-12], but there is a lack of standard that can be used to analyze the response of the various transducers in this particular framework. This kind of test, performed on different contact sensor, and the comparison of the parameters obtained from them suffers from lack of reproducibility as seen in chapter 1, since the human body is not a stable system on which a proper characterization can be performed [16].
For this reason the need of an apparatus that can mimic the biological system on which the sensor has to work arises, a device that gives two well-known outputs to a source signal: an acoustical signal and a vibration signal on a skin-like material.

The relationship between the vibration at the base of the neck and the acoustical vocal signal is hard to be accurately obtained. Both signals are related to the source of phonation, i.e. the air that flows through the glottis and drives the vocal folds into a self-oscillating motion, but the response of the two filters (tissue at the base of the neck and vocal tract) is not well reproducible due to various factors: tissue thickness, body fat, sensor position, attachment methods and sensitivity to acoustic stimuli. Furthermore it is hard to understand the features of the signal due to the characteristics of the measurand and to the artifacts introduced by the transducers

used to acquire the data.
The acoustic response of the vocal/respiratory system has been modelized [27], as the filtering effect which relates the voiced sounds radiated from the mouth to the neck surface acceleration [28], but requires a subject-specific calibration that can be only performed in a laboratory: it requires an oral airflow volume velocity measurements which requires a specific mask, and more in general is not a procedure that could be carried out outside of a medical clinic or a laboratory.
The research group activity involved on-field voice monitoring of teachers with the Voicecare, and usually the ad personam calibration procedure described in section 1.2.2 takes place in a normal room near the classroom where the teacher teaches his class. The device is also meant to be used in semianechoic rooms, call center or concert hall, to monitor the voice and prevent vocal pathologies, or in a voice support study framework which includes different test in different types of room.
For this reason, all the equipment that the measurement requires (calibration equipment included) needs to be easily transportable.
Against this background, the device needs a preliminary laboratory calibration, which consist in:

- performing a SPL calibration of the reference air microphone by means of a SPL calibrator, in order to obtain its calibration constant K_{mic} defined in equation 2.1;
- evaluating the frequency response of the contact microphone channel by means of a vibrant table [9] (which act as a reference for vibration) and including it in the form of a digital filter in the postprocess;

The use of a vibrant table may seem an easy way to evaluate the frequency response of a contact sensor in a comparison like the one described in section 2.1, but the vibrant table does not properly mimic the vibration damping effect of human tissues, and does not allow to consider the influence of the contact sensor geometry on the measurement performed on the human body.
The frequency response evaluation performed in [9] is still useful to lower the uncertainty in the SPL estimation from the skin vibration data, but sensors and devices needs to be tested and evaluated on something which mimics the response of the body part, on which the sensors and device are meant to be used.

In order to obviate to this problem, a similar system to test the vibration sensors to be used for recording lung sound has been already developed and tested [17, 18]. This device mimics the respiratory system and is able to give a well-known response to a certain stimula.

In this chapter, a similar apparatus developed on the basis of the phonatory system is described. This simulator acts as a stable generator that can make contact sensors characterization free from reproducibility problems.
Moreover, the proposed system allows other characteristic of the sensor to be estimated that cannot be obtained in vivo, such as the load effect of the sensor due to its dimension, shape and weight. The sensitivity to acoustical noise could also be easily investigated using an external sound source, avoiding the problems that have been faced during the in vivo measurements.

The development, construction and tests have been performed at the acoustic department of I.N.R.I.M.

2.2 System development

2.2.1 Design and prototype

The phonatory system is made up of the trachea, the larynx and the vocal tract. The larynx is the real source of phonation: it contains the vocal folds and the glottis. An air column ascends the trachea, an air channel that convolves the air that comes from the lungs in the larynx, and drives the vocal folds into a self-oscillation motion. In this way the glottis opens and closes itself and creates a pressure wave at the exit of the larynx. This vocal folds oscillation also creates the so called subglottal resonances, pressure standing waves in the trachea [19].

Nowadays the most commonly used instrument for the ambulatory monitoring of the vocal fold activity is the laryngostroboscopy, an optical device that allows to see the motion of the vocal folds during the phonation. It is the clinical gold standard for assessing the properties of the glottal phonatory function; it has a probe (which reaches the glottis via the nasal cavity) with a camera and a stroboscopic light, in

order to illuminate the glottis for a very narrow time window, and take a picture for every light pulse. The frequency of the pulse can be adjusted in order to capture all the phases of the opening-closing cycle of the glottis. It is a very useful instrument, that allows to directly see the vocal folds and their motion, so the physician can perform a visual verification of the vocal folds condition and motion and detect eventual disease, pathologies and incorrect behaviour.

Another instrument which has a similar purpose is the electroglottograph, an impedance meter that senses the relative impedance between the vocal folds and is able to acquire a signal that describes the opening-closing cycle of the glottis.
The spectral envelope of this signal is characterized by the greatest amplitude of the fundamental frequency, whilst the other harmonic components amplitude decreases as the frequency increases. A similar spectrum is also typical of the vibration signal acquired with a contact microphone at base of the neck during phonation, as seen in the contact sensor tests reported in chapter 1. The reason is that the vibration is mainly generated by the motion of the vocal folds, which is basically what the electroglottograph records.
In respect with laryngostroboscopy, nowadays electroglottography is no longer broadly used because it is more complicated to be operated. It is easier to perform a direct visual analysis of the motion of the vocal folds instead of reading a graph, which provides less informations.
In the spectral envelope of the acoustical voice signal, some of the higher harmonics are stronger than the fundamental frequency, due to the resonant effects of the vocal tract.
The primary pressure wave (the one which travel from the glottis towards the mouth) passes through the vocal tract, which act as a resonator: it changes the harmonic content of the original pressure wave by filtering some frequencies and amplifying some others (the so called formant frequencies). This particular frequencies differ in every voiced sound and depends on the oral cavity, lips and tongue disposition: different configurations correspond to different vowels, and each vowel is characterized by its own typical formant frequencies.

The aim of the project is to develop a system able to mimic the vocal apparatus (trachea, glottis/vocal cords, neck tissue and vocal tract) in order to obtain a test system with well-known relations between the vibration of a human-tissue-like

material and an acoustic signal. Some simulators that mimic the phonatory system have been already developed [20-22], as well as other airflow-drivern models (which use synthetic vocal folds or excised animal /human larynx) but in our case it is not required to develop an exact mechanical replica of the vocal apparatus, because the goal is the characterization of contact sensors in a stable vocal monitoring framework. So, it is not required that our system replicates the pressure wave produced near the glottis, or mimics the vocal folds movements, but it has to act like a "black box" with the two outputs previously mentioned.

The intention was to mimic the glottis and the trachea, because the vibration at the base of the neck is sustained by the motion of the folds themselves, but also by the resonances that occur in the trachea, which have their maximum at the termination (in our case, at the giugular notch). For our purposes a variable pressure system like the one used in [20-22] was too much, and it would been difficult to obtain a vibration on a skin-like surface. For this reason, a cylindrical plexiglass tube for the trachea and a speaker (with the same diameter of the tube) for the glottis have been chosen. The idea is that an emitting speaker, driven by the right signal, would simulate the vibration of the posterior part of the glottis. This vibration is the source of the one which is sensed in a vocal monitoring session, by means of a contact sensor placed at the base of the neck. So, the speaker had to be in contact with a skin-like and tissue-like material, placed in a window opened on the tube.

Tissue mimicking materials (TMM), or phantoms, are materials which show some mechanical, thermical or chemical characteristics similar to human tissues. They are widely used to test medical equipment, in order to quantify the effect and the possible damage of a particular medical device on the human body. They are also used to create models of parts of the human body to be used as practice instruments by the physicians. Medical students currently learn neck palpation by practicing on healthy, standardized patients; however, studies of similar procedures have shown that educational models with simulated pathology help to improve technique and confidence.
Different type of phantoms are create to mimic different types of human tissue, vessels and fluids, in order to test different types of medical equipment:

- ultrasound imaging systems exploits the backscattering of ultrasonic sound emission to create images of internal tissues (like in echography); this kind

of equipment are usually tested on tissue mimicking materials which had to exhibit some specific values of specific, acoustic-related parameters: backscattering, sound attenuation, sound speed and acoustic impedance [25, 26, 31, 33]. Moreover, focused ultrasound techniques are used to perform special surgery, which allows to operate on internal organs with less impact than the traditional surgery.
This particular kind of phantoms are prepared and tested at the I.N.R.I.M. in the acoustical physics laboratory.

- Radiation and electromagnetic imaging systems (like computed axial tomography and radiology) use the electromagnetical radiation backscattering and interference to obtain images of tissues, bones and vessel. This kind of radiation could drastically modifies human tissues and cause harm to the patients, so it is important to test this kind of instrumentation properly. The tissue properties that are important for the phantoms used in this framework are absorption and scattering coefficients, at different radiation frequency [30].

- Palpation model of various part of the body are widely used, as stated before [24, 32, 34]; in this case, the phantom materials had to exhibit mechanical characteristics similar to real ones, so the parameters to consider for this application are density and young modulus. However, it is important to point out that these characteristics and parameters are important for the general response of the phantom, which include the fact that a sensor placed on the body could alter the mechanical and acoustical response of the skin; for these reason they are still estimated for most materials.

For the project presented in this chapter, we need a material which exhibits the same "elasticity" of the neck tissues, so the parameter used for select the right material will be the Young's modulus, which is a mechanical property that measures the stiffness of a solid material. It defines the relationship between stress (force per unit area) and strain (proportional deformation) in a material in the linear elasticity regime of a uniaxial deformation.

The window on which the phantom will be placed has to cover a portion of the speaker, so that it could transmit some of the vibration to the tissue directly, like in the real phonatory system. Furthermore, the amplitude of the tube resonances

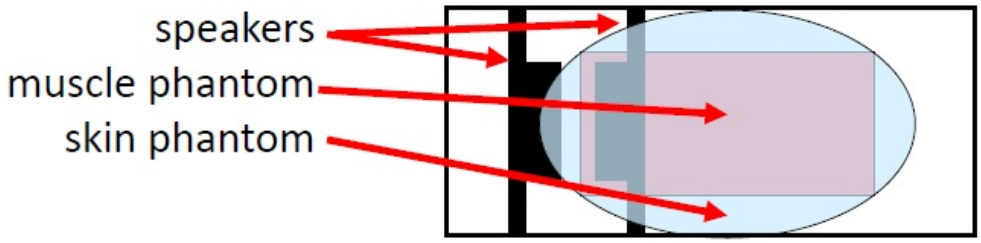


Fig. 2.1 Sensing zone scheme

are maximum at the closed end of it, so this particular window placement allows to maximize the vibration induced to the phantom. The window on the tube is the effective sensing zone, and it is the equivalent of the giugular notch, the place where the contact microphone is usually placed for a vocal monitoring session. In figure 2.1 two speakers are present: the reason will be explained later on.

The first prototype was assembled in order to test the positioning of the speaker and the window on the tube, which contains the phantom. A spare portion of tube was used, the definitive one would have been longer.

Two different types of tissue mimicking materials were used: one to mimic the internal tissues of the neck skin (dermis), and the other one to mimic the epidermis. The first one was a stiff Gellan Gum based hydrogel containing kieselghur and silicon carbide solid particles (to increment the elastic modulus), whose detail on preparation and characterization of its acoustic and mechanical properties can be found in [24-25]. The characterization has been made by the compression of a sample of the material and the measurement of its deformation; the density of the dermis is $\rho_{dermis} \simeq 1000\frac{kg}{m^3}$ and its Young modulus is $E_{dermis} \simeq 50kPa$; The phantom used to mimic it has $\rho_{TMMd} = 1034\frac{kg}{m^3}$ and $E_{TMMd} = 47 \div 57kPa$. A layer of this phantom, of the same thickness of the tube, was placed in the window previously mentioned.
A polyvinyl alcohol based phantom has been used as skin simulating tissue. Polyvinyl alcohol cryogel, (PVA-C), is a tissue-mimicking material, suitable for application in magnetic resonance (MR) imaging and ultrasound imaging. A 10% by weight polyvinyl alcohol in water solution was used to form PVA-C, which is solidified through a freeze–thaw process. Further details on its preparation and testing can be

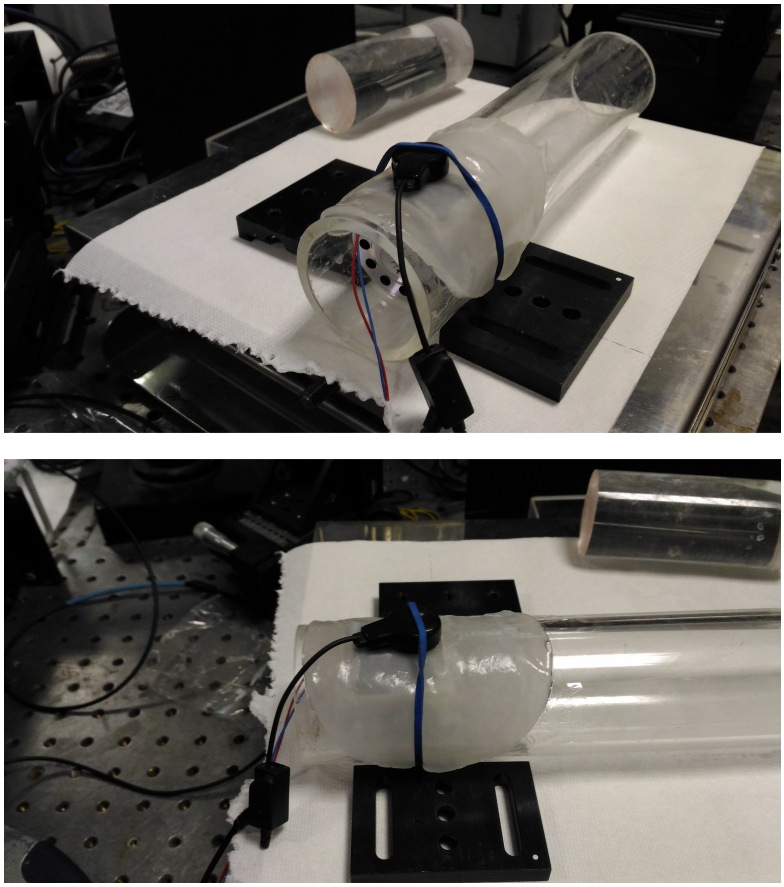


Fig. 2.2 Photos of the first prototype, with the ECM used in chapter 1 placed on the sensing zone.

found in [26]. A thin disk of it (3 mm, $\rho \simeq 1200\frac{kg}{m^3}$) was placed over the sensing zone, to cover the phantom which mimics the internal tissues.

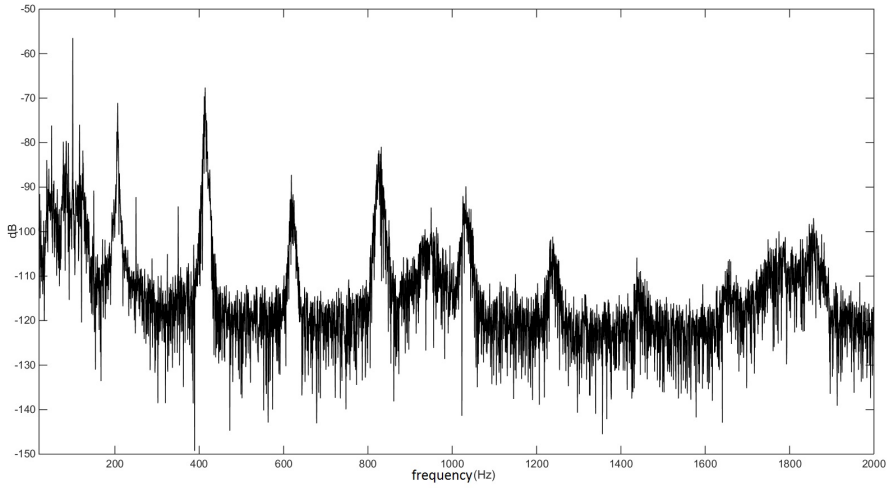


Fig. 2.3 Spectra of the vibration signal recorded during the test of the prototype, with sinusoidal source signal.

For this first test the source signal was a simple sinusoidal signal with 200 Hz frequency, created with Matlab. It was sent to the speaker by means of an amplifier, connected to a laptop.
The contact sensor used to test the prototype functioning was the bigger ECM used in the chapter 1, with the relative conditioning circuit. It has been attached to the simulator with a rubber band. Its signal has been acquired by means of a National Instruments USB 6211 acquisition board connected to the laptop, with 11 kHz sampling frequency. Data analysis has been performed using Matlab.
The spectra in figure 2.3 is obtained from the vibration signal acquired in the first test. This spectra is important because it is the evidence that the system can work: the vibration is well transmitted from the speaker to the phantom and it is strong enough to be sensed by the contact microphone placed on the sensing zone.

2.2.2 Definitive version

From now on, various contact sensors will be mentioned: they have been used to test and validate the system, in order to clarify if such system could be used to evaluate the behaviour of contact sensors to be used in a vocal monitoring framework. This

may sound recursive, but the only way to validate and test the output of the system is to record the vibration on the sensing zone. This could be done easily with contact sensors like the ones used in chapter 1 (a laser vibrometer will be used lately, but it is not a simple and easy-to-use system for sure). During the simulator functioning evaluation we will not investigate on the behaviour of the sensors in recording vocal material, but they will be used as simple vibration transducers: the attention will be focused on the differences between the signals recorded in vivo (on a human subject) and the signals recorded on the simulator. Lately, all the sensors will be tested on the simulator in order to briefly point out their behaviour.
The sensors are a Midland MIAE38 electret contact microphone (ECM), a Knowles BU-21771 accelerometer (ACC), a Meas-Spec CM-01B piezo film contact microphone (PMIC), and a commercial piezo throat microphone designed to be used as an external phone microphone (ARCH). The first two were already been used in the test described in chapter 1.

The PMIC is a contact microphone based on a thin film of polyvinylidene fluoride (PVDF) polymer: a polarized fluoropolymer which exhibits piezoelectricity. The sensor is composed by a rubber cylinder coupled with a disk of piezo film, a conditioning circuit (which provides preamplification to the weak signal produced by the piezofilm), all enclosed in a metal case.
The ARCH is a simple, classic piezoelectric ceramic sensor, equipped with a flexible plastic bow to keep the sensor in place. Unlike the other three sensors, the ARCH does not need the adhesive patch to stay in place.
After the tests on the prototype (that had confirmed the effectiveness of the approach) a new plexiglass tube of 55 cm of length was chosen: this choice was made to maintain the right proportion between the length of the "fake" trachea and the dimensions of the speaker.

The plexiglass tube is equipped with a moving removable end, in order to recreate the subglottal resonances by means of the closed/open tube modes. The simulator has three possible configurations: open end, closed-end and stopped-end. The tube length is 50 cm in the closed-end configuration and 55 cm in the open-end configuration. The stopped-end configuration is similar to the closed-end, but a disk of sound absorbing material is applied to the closed end of the simulator.

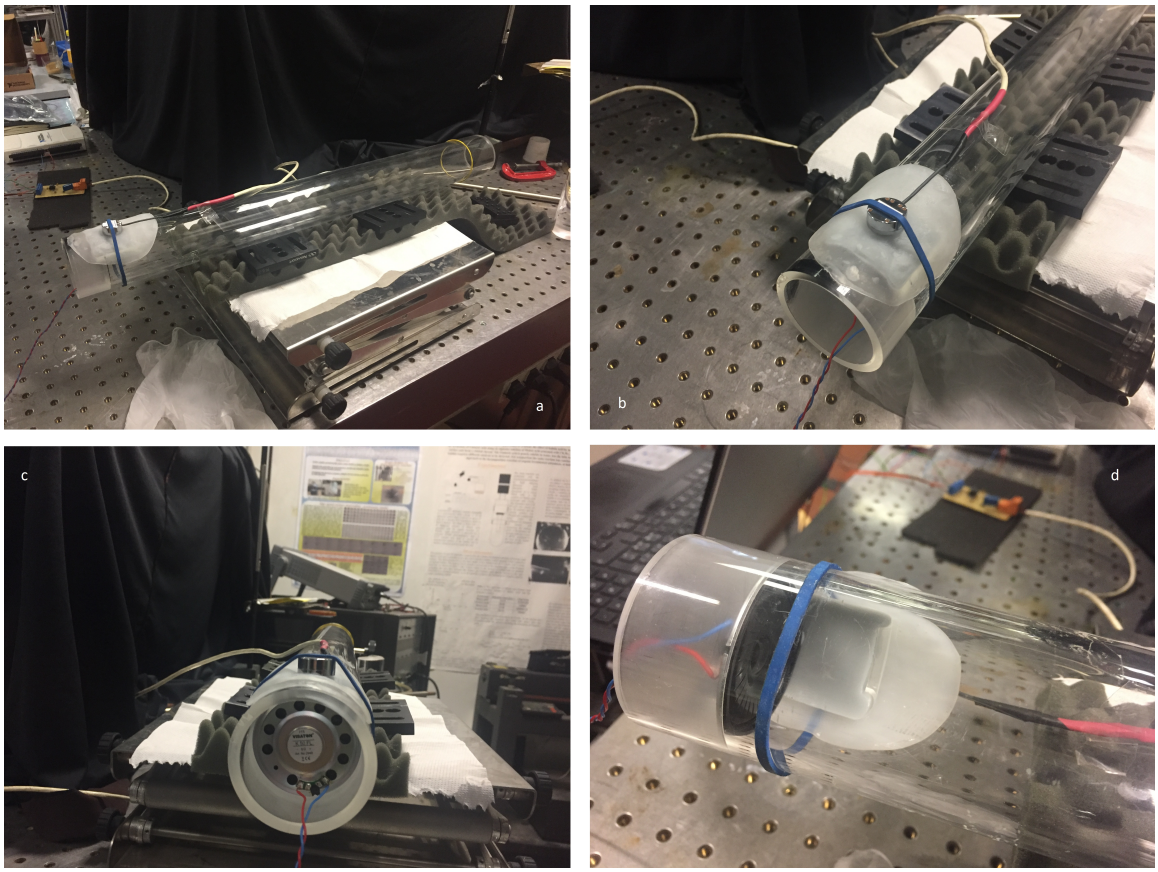


Fig. 2.4 Photos of the simulator with the new tube during the test: (a) the whole system, (b) detail of the sensing zone with the PMIC attached, (c) detail of the speaker and (d) the window with the tissue mimicking material used for neck tissues sees from below.

The issue was to find the right source signal to send to the speaker, in order to obtain a response as similar as possible to the real phonatory system. The solution was to use a real EGG signal, a signal recorded on human subject that describes the real motion of the vocal folds.

As stated above, the electroglottograph is a device used for the non-invasive measurement of the degree of contact between the vibrating vocal folds during voice production. The aspect of contact being measured by a typical EGG unit is considered to be the vocal folds contact area (VFCA). In order to measure VFCA, electrodes are applied on the surface of the neck so that the EGG records variations in the transverse electrical impedance of the larynx and nearby tissues by means of a small A/C electric current. This electrical impedance will vary slightly with the area of contact between the moist vocal folds during the segment of the glottal vibratory cycle in which the folds are in contact. However, no absolute measure of contact area is obtained, only the pattern of variation for a given subject, because the percentage variation in the neck impedance caused by vocal folds contact can be extremely small and varies considerably between subjects.

The EGG signal does not represent a real acoustic signal, so its use as a source signal sent to an acoustic speaker could seem improper, but no instrument able to record the pressure/acoustic emission directly above the glottis was available. One of the vibration signal acquired during the test on sensor could seem also appropriate as source signal, but it is affect by the influence of the sensor with which was acquired. The EGG is the only signal which is directly connected to vocal folds movements, and it is not filtered by the vocal tract or by the tissues of the neck, and it is not related to the contact sensor. It is affect by other measurement issues, but still its frequency content is very similar to the ones obtained for the contact microphones tested in chapter 1, so it appears to be the best choice, compared also to statistical noises or a simple sinusoidal signal.
It is also possible to obtain a glottal acoustic pressure signal from the phonation acoustic emission acquired with an air microphone, by means of an inverse filtering which removes the speech formants. This tecnique (glottal inverse filtering) could also produce a good source signal to be used in the phonatory system simulator.

Acquiring just the EGG signal was not enough: in order to validate the functioning

of the apparatus as phonatory system simulator, the vibration signal at the giugular notch had to be recorded too. In this way, the vibration output of the simulator would have been compared with the "real" vibration on a human body, produced by the same source signal used on the simulator. The acoustical signal (the voice) produced by the test subject during the EGG recording has been acquired too, for a reason that will be explained later.

A simultaneous recording of EGG, vibration signal at the base of the neck and voice emission were performed on a human subject (male, emitting the vowel /a/, voice recorded at distance of 15 cm, on axis). The acquisition took place at the Molinette hospital of Torino (otolaryngology ambulatory). The spectra obtained from the signals acquired on the human subject will be called “in vivo spectra”. The vibration and the acoustic emission were recorded by using the same measurement chain used in chapter 1 to evaluate the contact sensors.
By comparing the in vivo recording with the recording done on the simulator (acquired with the same contact sensor), the characteristics of the sensor itself are not important anymore, because they affect the two acquisitions in the same way.

Unfortunately, the electroglottograph was not available to carry out an extensive data acquisition session on different subjects. Although, it could be useful to eventually create a database of simultaneous measurements off EGG, acoustic signal and vibration at the base of the neck signal during phonation, acquired from subjects of different gender, age and constitution. It will be useful to compare the impact of different anatomy on the acquired signals, and possibly create different digital source signals to be used on the simulator, tuneable and adjustable to reproduce the phonation of different individuals.

Various tests were performed to validate the functioning of the apparatus and the measurement chain, but the tests used for the real validation of the complete simulator will be reported in section 2.3.

In figure 2.5 and 2.6 the spectra obtained in the first test with the EGG are reported. In this test, the EGG was used as source signal sent to the speaker, and the vibration was recorded with the PMIC. The most important result of this test, besides the proof that the system and the measurement chain work, is that the shape of the vibration spectra is very similar to the one obtained on human subjects, and so

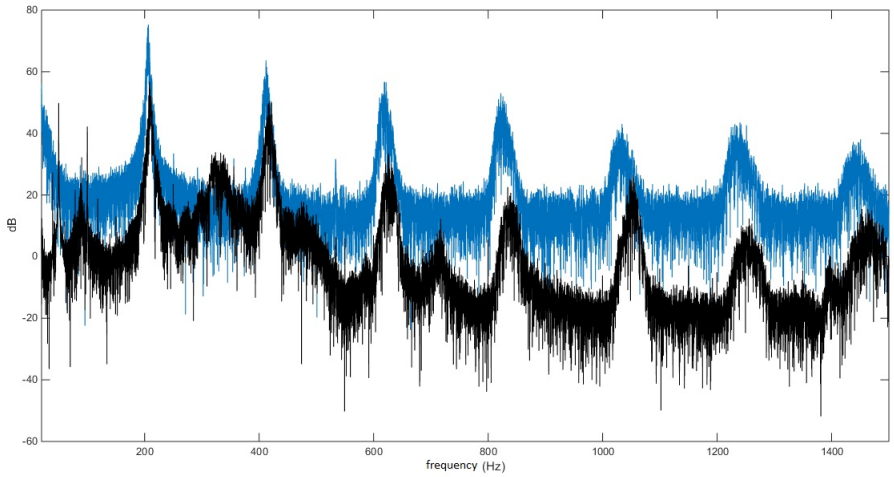


Fig. 2.5 Spectra of the EGG signal (blue) used as a source signal, and spectra of the vibration signal (black) acquired with the PMIC on the simulator driven by the EGG signal. The simulator was in the closed-end configuration.

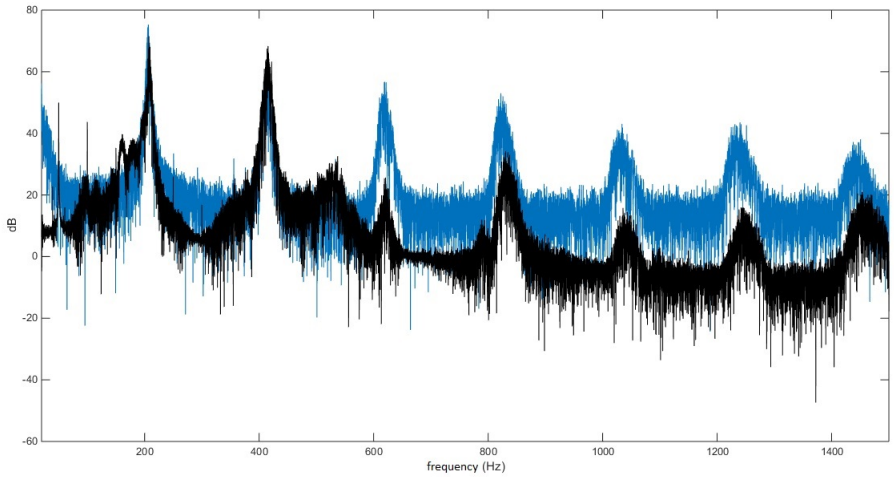


Fig. 2.6 Spectra of the EGG signal (blue) used as a source signal, and spectra of the vibration signal (black) acquired with the PMIC on the simulator driven by the EGG signal. The simulator was in the open-end configuration.

that the simulator can mimic the real phonatory system.
This test was performed by using the PMIC; it was just a preliminary test, performed in order to obtain some quantitative information on the simulator behaviour. Qualitative information obtained from more extensive tests, described in section 2.3.2.

2.2.3 Vocal tract simulator

The aim of the simulator is to mimic the real phonatory system, in order to give a reference that can be used to test contact microphones. All the vocal dosimeters that use that kind of sensor are also equipped with a normal air microphone, in order to perform a calibration of the contact channel.
For this reason, the simulator needs to have an acoustic output, which can mimic the final tract of the phonatory system, the vocal tract, and can act as a filter on the source signal emitted by the fake glottis (the speaker). In this way, we could obtain a complete test system that produces a vibration and an acoustic output originated from the same source signal. The relation between these two elements is supposed to be stable, so that the issues due to the tests on human subjects emerged in the chapter 1 could be overtaken.

But how could we mimic the vocal tract? The answer can be found in [23]: Vampola et al. performed a modelization of a vocal tract while emitting various vowels. The modelization was performed by mixing different techniques: MRI scan, acoustic modelization and finite element modelization. The result is a linear chamber (pseudo 1D model) which exhibits the same resonances and filtering effect of a real vocal tract while emitting various vowels. For the simulator we used the model related to the /a/ vowel, presented in figure 2.7 (up).

The starting point was the profile of the linear model, used to prepare a mesh (3D model) with Autocad to be 3D printed. The result is presented in figure 2.8.

Another speaker, identical to the one already installed on the simulator, was placed near the first one but contrariwise oriented: it had to emit in the opposite direction, inside the resonator. It was connected in parallel with the first one, so both

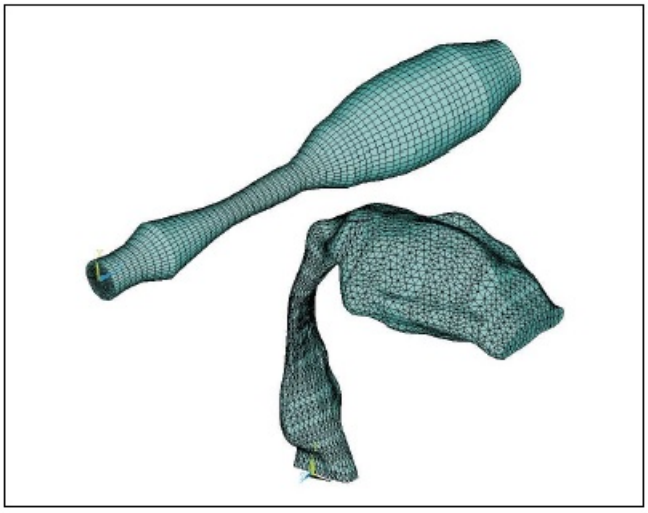


Fig. 2.7 Mesh of the vocal tract model for the /a/vowel from Vampola et al.: pseudo 1D model (up) and 3D Finite Element Model (down).

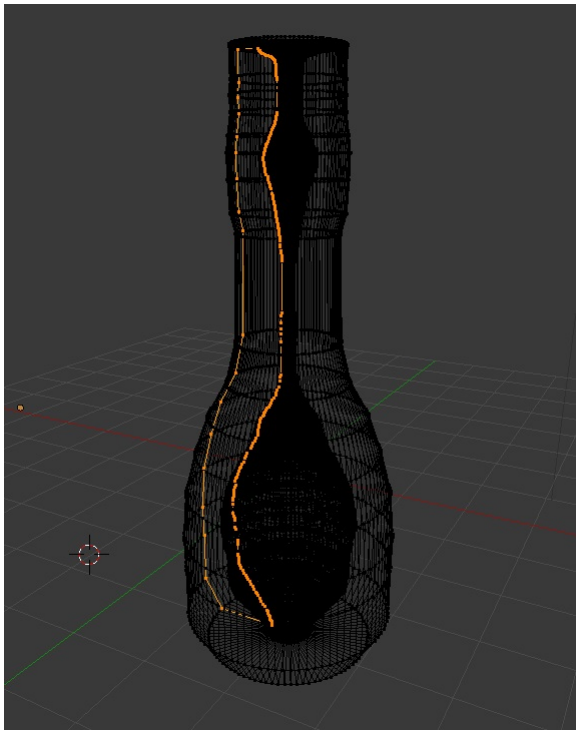


Fig. 2.8 Mesh of the vocal tract model for the /a/vowel to be 3D printed.

of them can emit the same signal at the same time.
The simulator is now complete: it has a vibration and an acoustic output. The acoustic output is acquired by means of the same air microphone and circuit used in the forst chapter, in the contact microphones tests. Photos of the complete system (simulator and sensors) are presented in figure 2.9.

Tests have been made in order to point out the behaviour of the resonator; spectra and signals obtained in closed-end configuration are presented in figure 2.10. The used source signal was the EGG, and the graphs are referred to the acquisition made with the PMIC in the closed-end and open-end configuration. The graphs related to the other contact sensors are presented in Appendix 2.

In the closed-end configuration the signal exhibits frequency components higher than the fundamental frequency, due to the fact that the closed-tube resonances are stronger and more energetic than the open-tube ones.
In the open-end configuration, the tube resonances do not affect the acoustic output as much as the vibration output.
The spectra presented in figure 2.10 (c) explains quite well how the resonator works: it filters the acoustic signal that comes out from the speaker (which can be assumed similar to the vibration signal in terms of frequency components) by amplifying some frequencies and attenuating some others, and creating the so called "formant frequencies".
All these tests have been made in order to obtain some qualitative information on the simulator functioning. The quantitative results will be presented in section 2.3.

2.2.4 Load effect measurement

During the first tests, the gellan gum-based hydrogel used to mimic the internal neck tissues introduced some issues. Firstly, it did not show great time stability: it lost the requested mechanical characteristics within two weeks from its preparation. Secondly, the preparation itself is complicated: the gel needs to be dripped in a closed space and then sealed, or the surface which remains in contact with the open air will dry too much. So it is very difficult to create a well-made curved layer of gel inside the window on the tube.

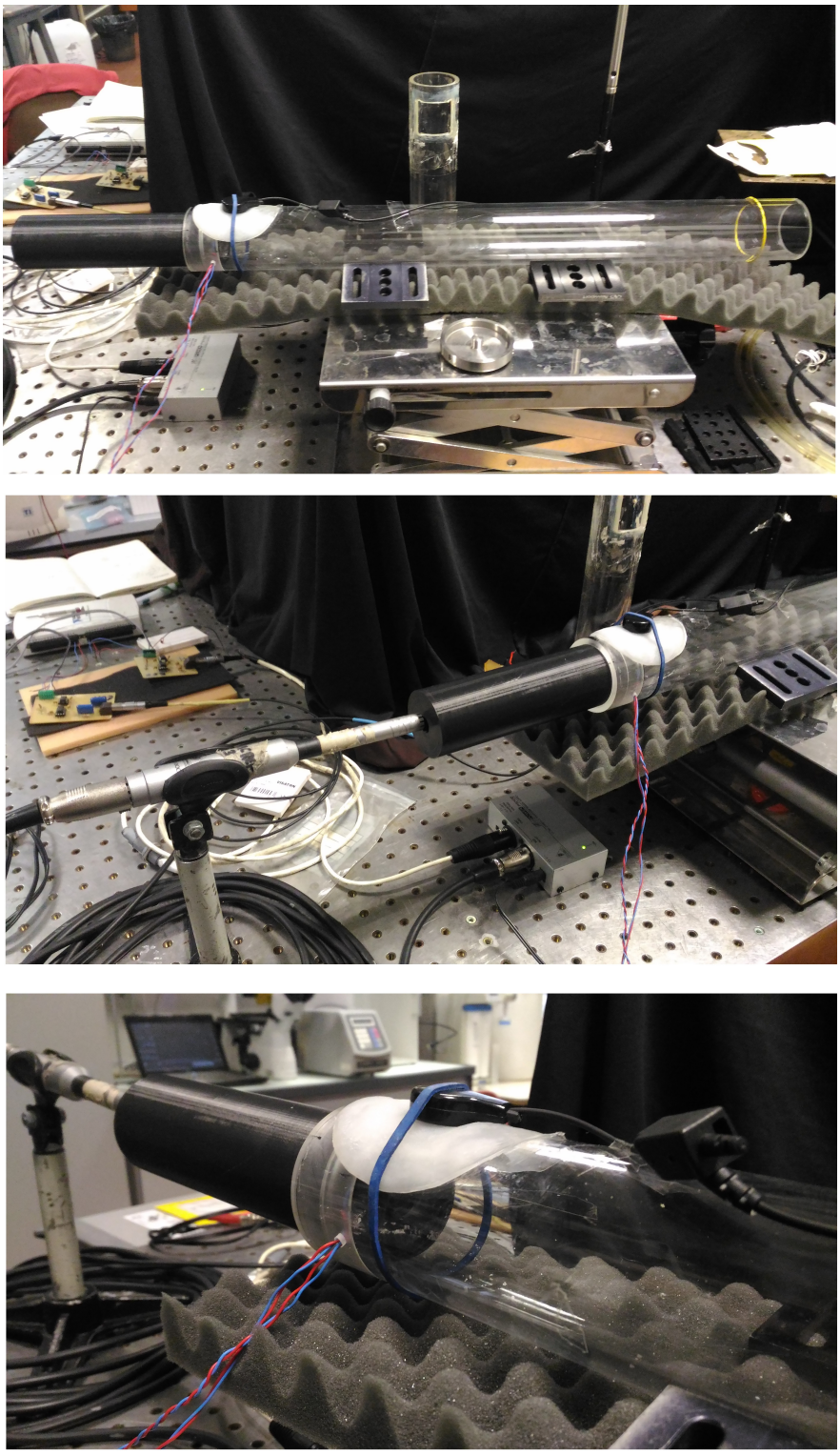


Fig. 2.9 Phonatory system simulator equipped with the vocal tract resonator and sensors (contact microphone and air microphone).

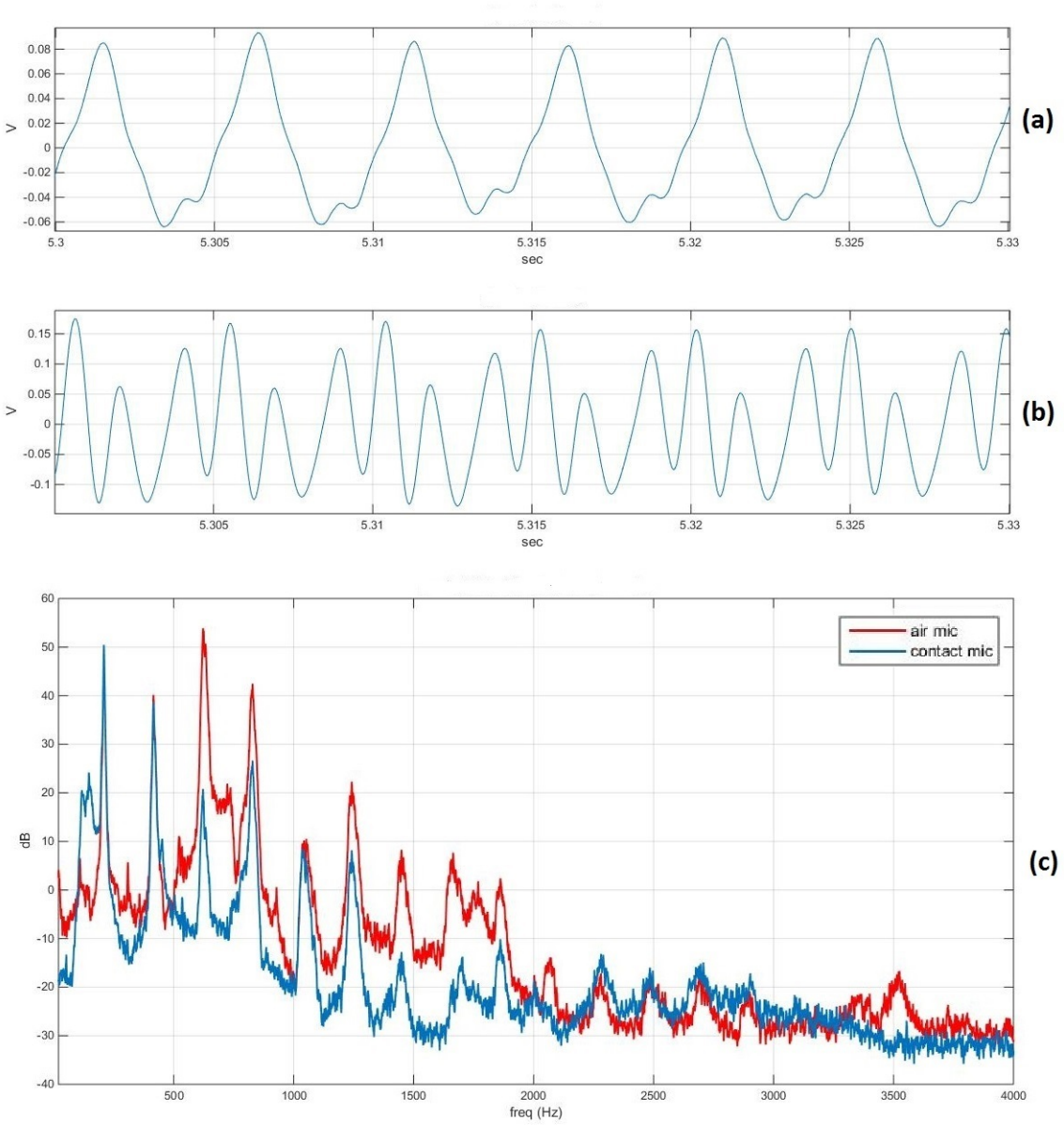


Fig. 2.10 Vibration signal (a), acoustic signal (b) and spectra (c) obtained during the first test of the complete simulator. Acquisition made by means of the PMIC, simulator in open-end configuration.

For these reasons, a latex rubber-based material has been used to mimic the neck internal tissues. The mechanical properties of the latex rubber-based tissue used in this application are stable in time (less degradability) and similar to gellan gum-based hydrogels. Furthermore, the preparation of the layer with this material is much simpler: the liquid latex solidifies at not so high temperatures, and it does not need to be sealed.

The preparation consists in:

- closing the concave side of the window with a thin plastic sheet;
- dripping a thin layer of liquid latex (about 2 mm);
- blowing on it with a hot air source;
- repeating

This procedure of creating one thin layer of latex on another allows to easily embed something in it. When the PMIC was bought, some other piezo film sensors were bought too. They consist in a thin layer of a piezoelectric material (with two electrodes connected to the two sides of it) that generates a current signal proportional to its geometric modifications, and can be used as a vibration transducer. It has been tested as contact sensor for acquiring the vibration at the giugular notch, because it is very thin and could hypothetically couple well with organic tissues, but it was too difficult to attach to the neck skin.

During the first tests on human subjects, arose the idea of considering the influence of the sensor of the human phonatory system as a useful parameter to evaluate the sensors to be used in this particular framework. For example, a big sensor could press too much on the giugular notch and could affect the way the tissues vibrate. This can be called the load effect of the sensor. Generally speaking, the load effect can be defined as an estimation of how much a particular object used as a measurement instrument influences and disturbs the system under measurement, a difference between the operation of the system with and without the measuring instrument. In our case, this means that the vibration at the base of the neck should be recorded with and without the sensor attached.
In the tests on human subjects, this can not be done because it is impossible to record that particular vibration without a contact sensor. But on a simulator there may be

a way to accomplish that. The first idea was to try with a laser vibrometer, which can record a vibration on a surface without contact. But another problem arose: in order to make the two measurements (with and without sensor) comparable, they have to be performed with the same instrument. The tests pointed out that this was not possible.

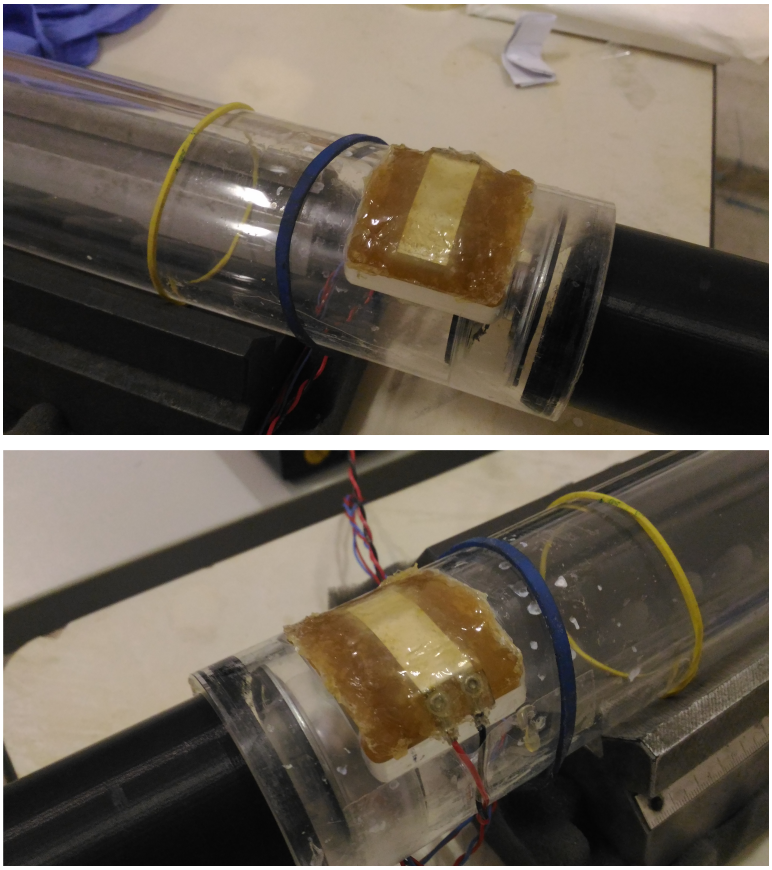


Fig. 2.11 The piezofilm stripe embedded in the latex rubber based TMM.

With the change of the TMM used for the skin internal tissues, it was possible to embed a sensor into it. This sensor can acquire the vibration directly, without any other external measurement system. The piezofilm stripe is perfect for this application, and allows to compare the vibration of the TMM in the unloaded simulator (no sensor attached to the sensing zone) with the vibration obtained without the sensor attached, to obtain an estimation of the load effect. The stripe does not affect the measurements because of its slight thickness and its high flexibility, which do not alter the mechanical response of the TMM.
The used acquisition board is multichannel, so the three signals (air microphone,

contact microphone and piezofilm) can be acquired simultaneously.

2.3 Experimental tests and results

Specific tests were carried out to characterize the simulator frequency behaviour, to validate its capability to mimic the phonatory system and to test the four contact sensors previously mentioned: a Midland MIAE38 electret contact microphone (ECM), a Meas-Spec CM-01B piezo film contact microphone (PMIC), a Knowles BU-21771 accelerometer (ACC) and a commercial piezo throat microphone (ARCH). Every sensor was equipped with an analog signal conditioning circuit specifically designed, enclosed in a modular system to make the connection switch easier.
The acoustic signals have been sensed with a Behringer ECM2000 microphone, the same used for the tests reported in the first chapter; this microphone can be coupled to a standard sound level calibrator to ensure metrological traceability.
All signals have been acquired by means of the National Instruments USB 6211 acquisition board connected to a PC, with 11 kHz sampling frequency. Data analysis were performed with Matlab software, using ad-hoc scripts to perform calculations, FFT and digital filtering.

The definitive version of the simulator on which the tests have been run consists of:

- trachea: hollow plexiglass tube with a diameter of 5 cm;
- glottis / vocal folds / source signal: two speakers that emit in two different directions;
- vocal tract: a 3D-printed model of Human Vocal Tract (HVT);
- neck tissues: tissue-mimicking phantom materials (skin and muscles).

The two speakers recreate the glottis activity by emitting two pressure signals: one signal propagates in the plexiglass tube (trachea) and the other one in the vocal tract simulator.
The plexiglass tube used to simulate the trachea is 55 cm long and it is equipped with a moving removable end, in order to recreate the subglottal resonances by means of

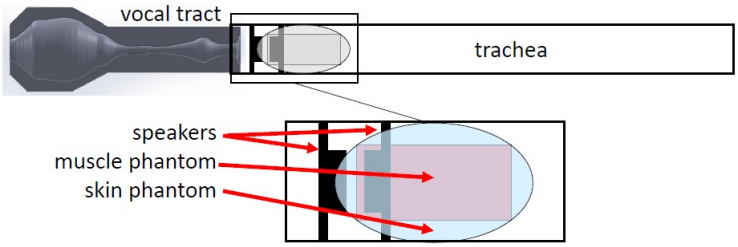


Fig. 2.12 Scheme of the definitive version of the simulator

the closed/open tube modes.
The simulator has three possible configurations, previously described: open-end, closed-end and stopped-end.
The vocal tract is a 3D-printed hollow resonator, which is based on a model of the vocal tract while emitting a vowel /a/, as proposed by Švec et al. This resonator allows the formant frequencies from the pressure signal originated from the speaker to be amplified.
The muscles and the internal tissues of the throat have been mimicked by using a latex rubber-based Tissue Mimicking Material (TMM). A polyvinyl alcohol based TMM thin disk has been used as skin simulating tissue.
In order to obtain information on the load effect of the tested sensors, a piezofilm stripe was embedded in the latex rubber TMM.
The speakers of the simulator are driven by an audio amplifier connected to the audio output of the PC, which is used to generate the source signal.
The measuring chain is presented in figure 2.13. The tests were performed at INRIM acoustic department, in an acoustically isolated room.
From now on, we refer to the phonatory system simulator as the TS (which stands for Throat Simulator, the first name used for the apparatus).

The performed tests were:

- **simulator frequency characterization**: the TS is driven by a linear frequency sweep sine wave (15 seconds duration, 20 Hz ÷ 5 kHz), the phantoms vibration is recorded by means of a laser vibrometer to obtain the simulator frequency response;
- **effectiveness evaluation**: the spectra of the signals (vibration and sound emission) obtained on a human subject emitting an /a/ vowel are compared to

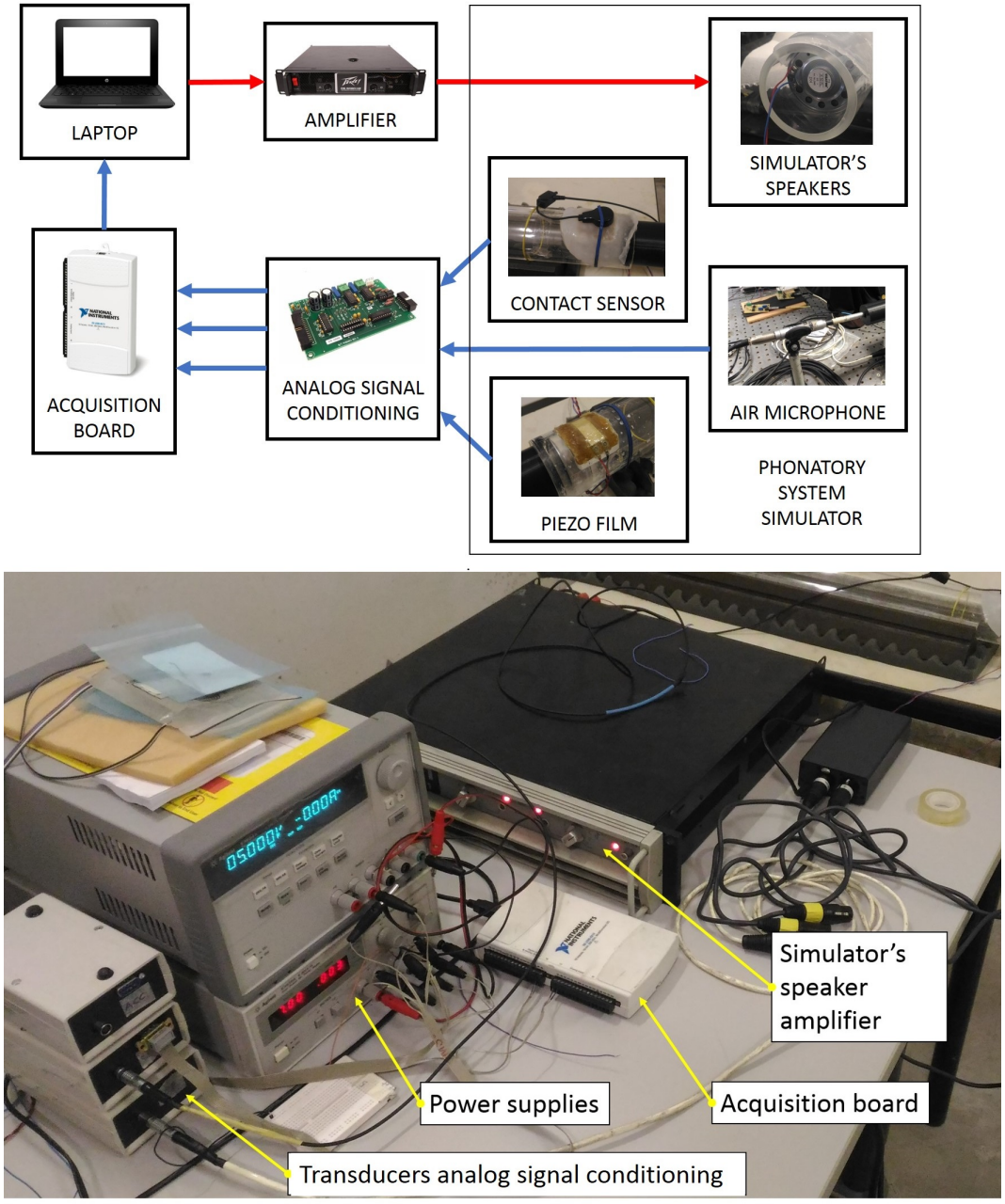


Fig. 2.13 Measuring chain used for the tests on the definitive version of the simulator: it conditions the signal in order to be right acquired by the acquisition board, provides the power supply to the analog conditioning circuits and to the sensors that require it, drives the speakers of the simulator and acquires the signal of the air microphone, the contact microphone and the piezofilm.

those acquired on the simulator, which is driven by the EGG signal recorded during the in vivo acquisition;

- **repeatability**: the TS is driven by the linear frequency sweep sine wave. The same measure is repeated 3 times in a slightly different conditions, and the spectral standard deviations among the obtained spectra are calculated;
- **sensors load effect**: the signals acquired by the embedded piezofilm are used to evaluate the load effect of the 4 sensors previously mentioned. The piezofilm spectra obtained with the unloaded TS (no contact sensor attached) are compared to the spectra referred to the TS with the sensors attached to the sensing zone;
- **sensors characterization**: the 4 sensors are tested on the simulator by using a linear frequency sweep sine wave as a source signal.

2.3.1 Simulator frequency characterization

The apparatus has been characterized by means of a Laser vibrometer (Polytec CLV-2534), in order to obtain the system vibration response with no mechanical sensor attached to the phantom.
The simulator has been placed on the optical bench and the laser was focused on a piece of reflective material placed on the skin-mimicking phantom. A linear frequency sweep (15 seconds duration, 20 Hz $\div$ 5 kHz) sine wave was used as a source signal.

The spectra presented in figure 2.15 refer to the acceleration signals obtained with the vibrometer, for the three configurations of the TS. The peaks represent the acoustical resonances of the speakers and the tube in the three configurations. These spectral envelopes define the filter effect of the simulator, which includes the electrical-to-acoustical response of the speakers, the effect of the tube resonances and antiresonances and the attenuation of the phantoms. This characterization can be used to weight the response of the sensor tested on the TS.

From the graph emerges that the closed-end and stopped-end envelopes are quite similar, except for the harmonics over 2 kHz, that in the stopped-end configuration

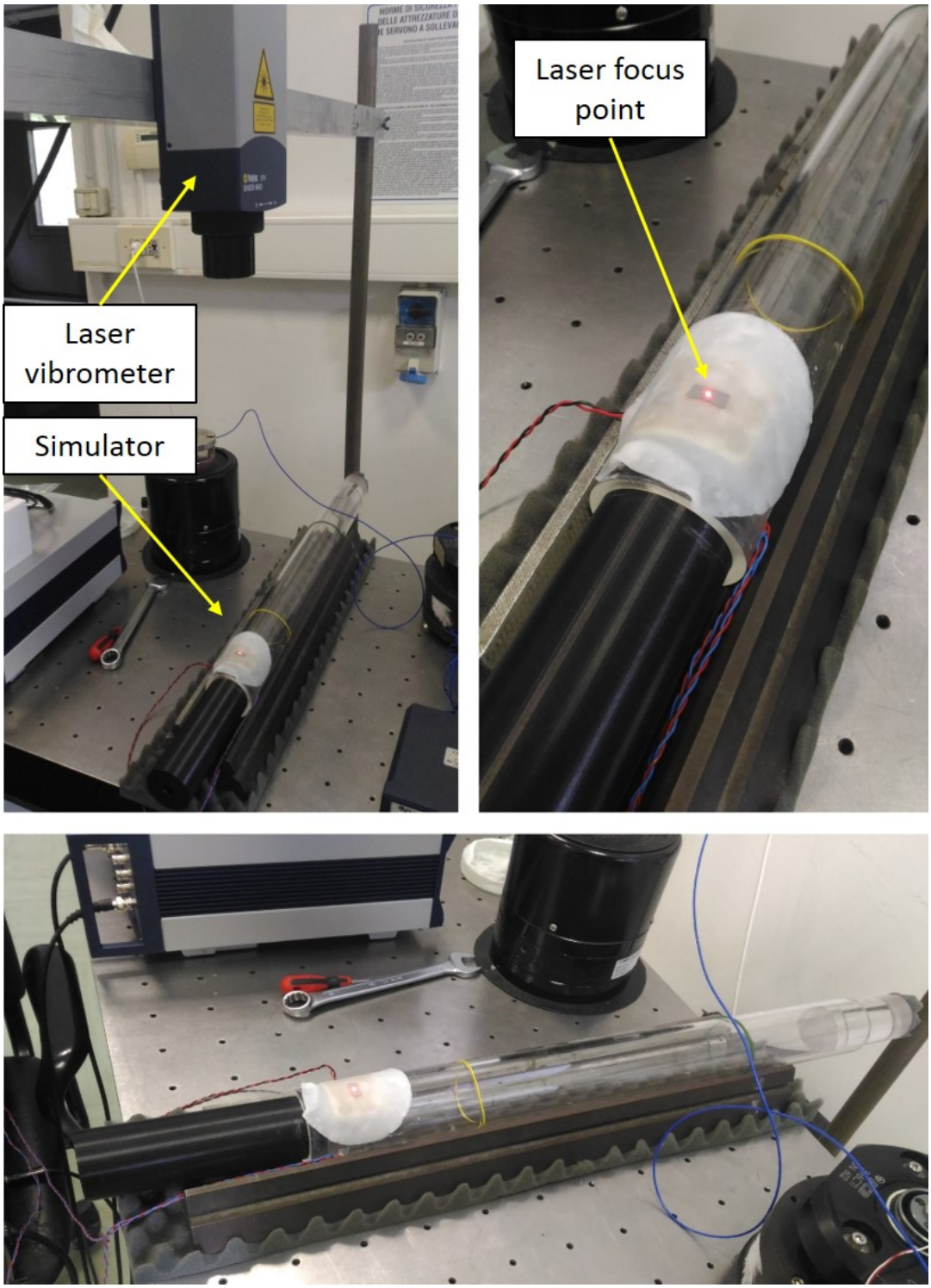


Fig. 2.14 Laser vibrometer measurement setup.

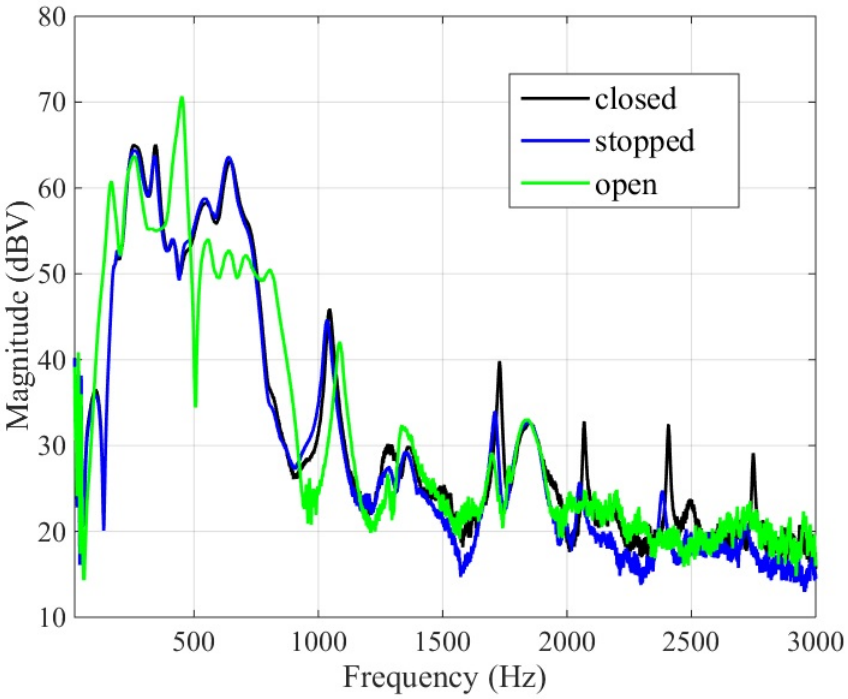


Fig. 2.15 TS characterization obtained with the laser vibrometer for the three configurations.

are attenuated by the used sound-absorbing material.
The peaks change their position from the open-end to the closed-end configuration due to the different resonance frequencies, except for the two peaks at 1280 Hz and 1880 Hz. They are common to the three envelopes and they probably are the loudspeakers resonances.
In this measurement the noise floor can be estimated around 20 dBV.

2.3.2 Effectiveness evaluation

The aim of this test is to evaluate the effectiveness of the simulator to mimic the vocal apparatus. In order to perform this evaluation it is necessary to obtain the response of a real vocal apparatus recorded with the same sensors and measurement chain used for the simulator. This has to be used as a reference: such response can be compared to those obtained on the simulator, to evaluate how well it mimics the real apparatus.

A simultaneous recording of EGG, vibration signal at the base of the neck and

voice emission was performed on a human subject, as mentioned in the previous section. This recording has taken place at the otolaryngology ambulatory of Molinette hospital of Torino, by using the same measurement chain previously described.

The EGG signal has been used as a source signal for the simulator, the same source that generated the phonation used as a reference for the validation. All the in vivo tests were performed three times, with three of the four considered sensors. The ARCH was not used in this tests because of its shape, which did not permit the contemporary acquisition of the EGG and the vibration signal.

Afterwards, tests on the simulator were carried out by recording the acoustic emission, the vibration and the piezofilm signal, for three of the contact sensors and for every simulator configurations, with the EGG as source signal sent to the speakers. Every test has been repeated 3 times in slightly different conditions, for each sensor and configuration, and the average spectra of the three repetitions have been calculated. The repeated measurements have been useful to obtain an information on the repeatability of the procedure.

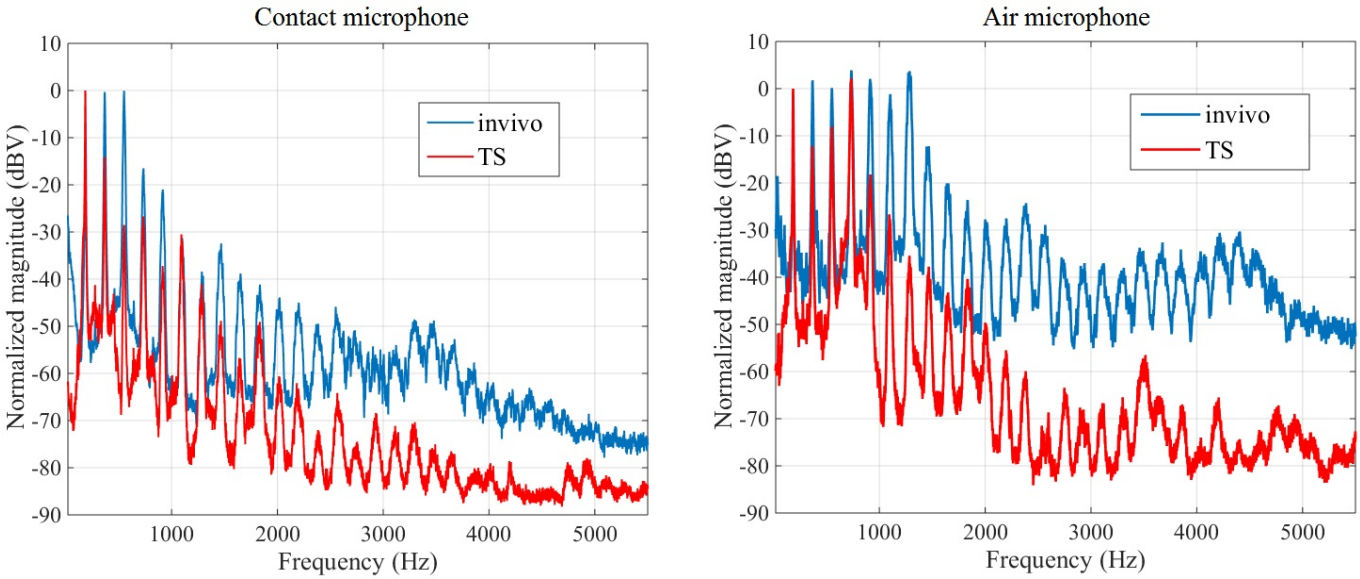


Fig. 2.16 Comparison between in vivo and TS vibration signal spectra (left) and between in vivo and TS acoustic signal spectra (right). Measurements done with ECM, TS in open-end configuration.

At this point, for each contact sensor, 4 spectra are obtained:

Table 2.1 AVERAGE SPECTRAL DIFFERENCES BETWEEN TS SPECTRA AND IN VIVO SPECTRA. The first column indicates the contact microphone used for the measurement.

		average spectral difference (dBV)	
	TS configuration	***contact microphone***	***air microphone***
ACC	open	10.3	29.4
	closed	14.2	28.0
	stopped	13.9	28.1
ECM	open	11.8	24.1
	closed	6.8	23.9
	stopped	7.9	23.9
PMIC	open	15.1	26.8
	closed	11.5	28.0
	stopped	11.4	28.3

- the in vivo spectra of the air microphone signal;
- the in vivo spectra of the contact sensor signal;
- the TS spectra of the air microphone signal;
- the TS spectra of the contact sensor signal.

All these spectra are obtained from "the same source signal": the EGG recorded during the in vivo acquiring session.

All the spectra have been normalized to the magnitude of the first harmonic, and the differences between the TS average spectra and the in vivo spectra have been computed. This has been done in order to obtain information on how the TS behaviour is different from the one of a real phonatory system.
Hereafter, the averages of the absolute values of these differences were calculated, for the air microphone signal, among the 20÷5500 Hz frequency interval. For the vibration signals these averages were calculated in the 50÷3000 Hz frequency interval, because from the in vivo tests performed for this thesis, emerges that the vibration at the base of the neck related to the phonation does not present significant components above 3000 Hz. Results are summarized in table 2.1.

Figure 2.16 shows the spectra of the vibration signal and the sound emission acquired on the TS in the open-end configuration compared to the spectra related to the in vivo measurement. Referring to the contact microphone (ECM in this case), the TS spectra appears to be quite similar to the in vivo one. Moreover, the averages of the TS-in vivo spectral differences are lower than 15% of the higher peak amplitude. These facts confirm the effectiveness of the simulator.

For what concerns the air microphone spectra, the resonator creates the formant frequencies of the /a/ vowel. In this case, the differences among the two spectra (figure 2.16 right) and the resulting values of the spectra average differences are considerable because the resonator is not designed on the bases of the vocal tract of the person who has generated the in vivo signals. This does not compromise the effectiveness of the TS because the aim is not to reproduce a particular vocal apparatus, but to have a repeatable response and a stable vibration-to-acoustic transfer function. The spectra for all the considered sensors and configurations are reported in Appendix B.

Some qualitative consideration can be done on the time-domain waveforms, reported in figures 2.17 and 2.18; the signals recorded in vivo are compared to the ones recorded on the simulator. The air microphone signals confirm what already state above about the big difference between the simulator and the real body. The contact microphone signals confirms what already stated too, because some of the harmonics present in the in vivo signal are less evident in the simulator signal, but the signals are quite similar.

Unfortunately, the values reported in table 2.1 do not help in selecting the best TS configuration, because the lower value for the average spectral difference occurs in a different configuration for every contact microphone.

2.3.3 Repeatability

The main goal of the simulator development was to create an apparatus on which a repeatable measurement can be performed. A test has been performed in order to exploit the repeatability of the measurements done on the simulator.

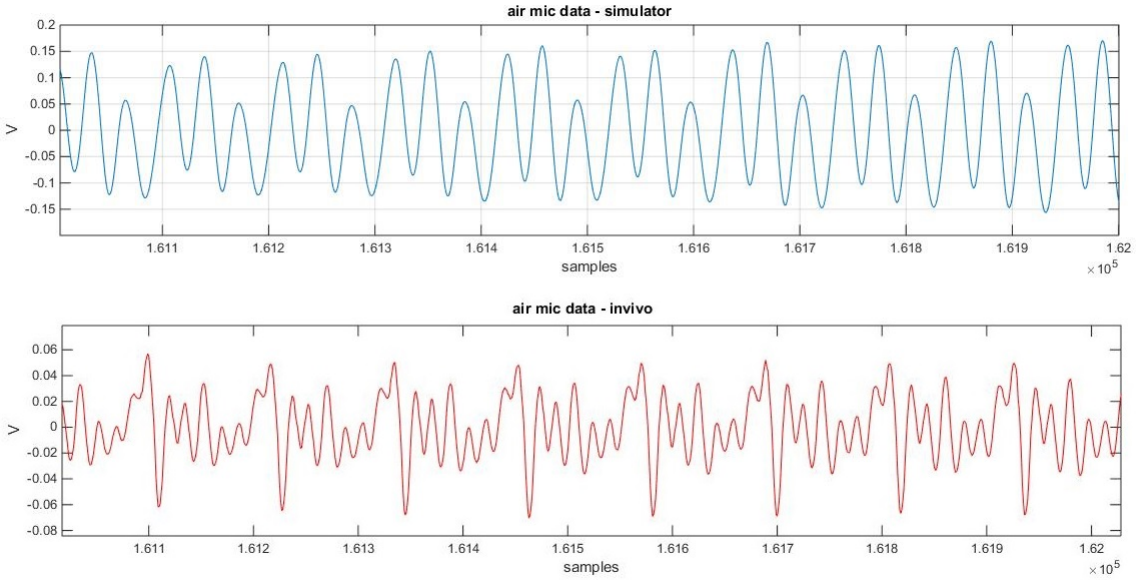


Fig. 2.17 Comparison between in vivo and TS air microphone signal. TS in open-end configuration, PMIC attached to the sensing zone.

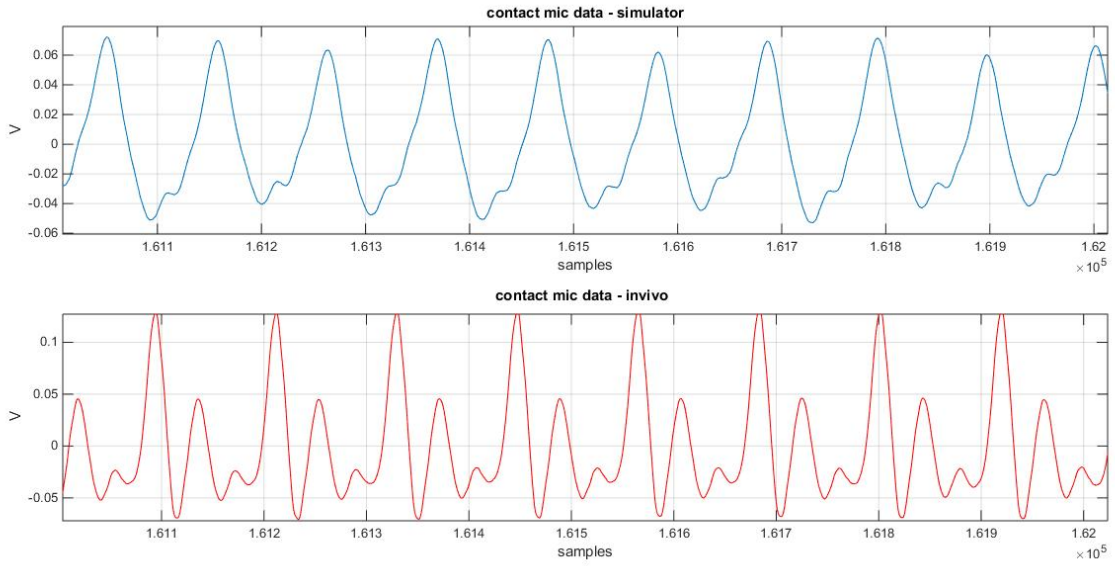


Fig. 2.18 Comparison between in vivo and TS vibration signal. Measurements done with PMIC, TS in open-end configuration.

A frequency linear sweep sine wave (15 seconds duration, 20÷5500 Hz) was used as source signal to drive the TS, and three outputs were recorded: the sound emission with the air microphone, the vibration by means of the contact sensor and the piezofilm signal. In this case all the sensors, the microphone and the piezofilm were used as instruments to prove the repeatability of the measurement done on the TS. This test was performed three times for each sensor and for every simulator configuration (open-end, closed-end, stopped-end).

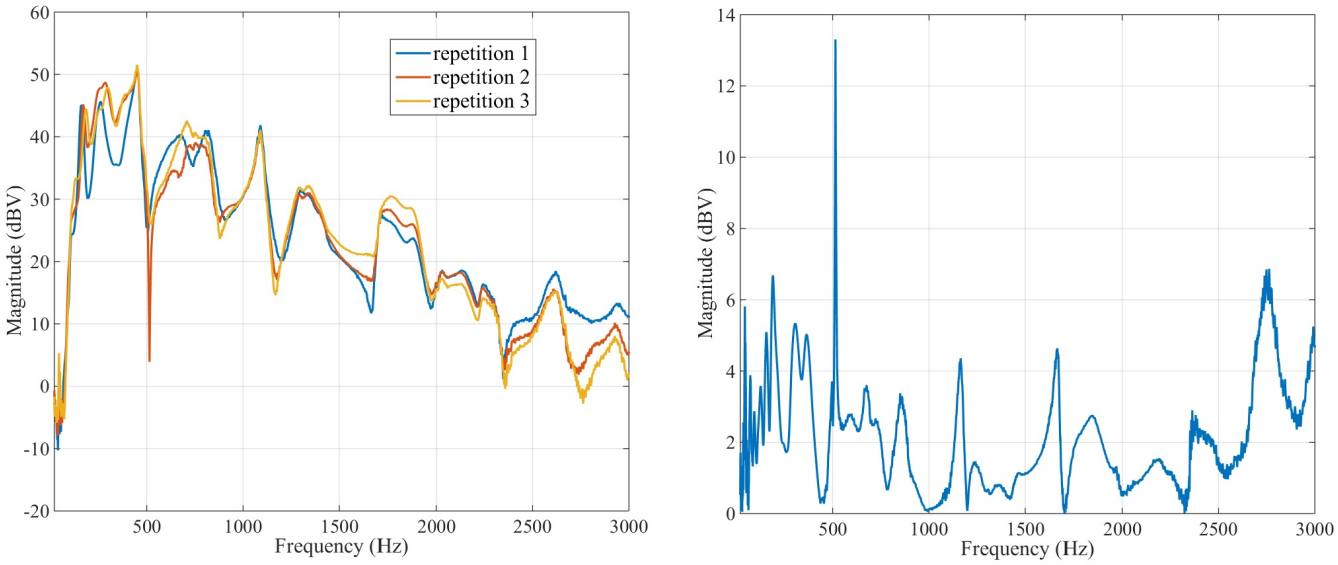


Fig. 2.19 Contact microphone channel spectra of the three repetitions of the same test (left), spectral standard deviation among the three repetitions (right). Measurements done with ECM on the TS in open-end configuration.

The FFT of every recorded signal has been calculated, and the three spectra of the repetitions of the same measurement were compared to obtain a standard deviation for every spectral point. So, for every contact sensor - TS configuration combination, a "spectral standard deviation" for every channel (contact sensor, air microphone and piezofilm) has been obtained.

The average of this spectral standard deviation among the 20÷5500 Hz frequency interval has been calculated, in order to have an indicator of the measurements repeatability.
Examples of the spectra of the vibration signal of three different repetitions are

Table 2.2 AVERAGE SPECTRAL STANDARD DEVIATIONS AMONG THREE REPETITIONS OF THE SAME MEASUREMENT, FOR EACH OUTPUT IN EACH TS CONFIGURATION. The contact sensor used for every measurement is reported in the first column.

		average spectral standard deviation (dBV)		
	TS configuration	***contact microphone***	***air microphone***	***piezofilm***
ACC	closed	3.4	1.1	2.2
	open	3.1	0.7	1.8
	stopped	3.4	1.1	2.6
ARCH	closed	4.7	1.6	3.0
	open	4.2	0.7	2.4
	stopped	4.4	1.5	3.1
ECM	closed	2.7	0.6	2.4
	open	1.9	0.4	2.2
	stopped	2.4	1.0	2.2
PMIC	closed	2.7	0.9	2.0
	open	2.9	0.6	1.5
	stopped	4.7	2.9	3.6

presented in figure 2.19 (left); it refers to the ECM used on the simulator in the open-end configuration. The spectral standard deviation obtained from these repetitions is presented in figure 2.19 (right). This procedure has been applied to the air microphone and the piezofilm recorded signals too.

Moreover, the differences between the average spectral standard deviation values for the various contact microphones give preliminary information on how the different contact sensors affect the system under measurement; the sensor/configuration combinations which exhibit the higher average standard deviation values for the contact mic channel present high values in the other two channels too.
This could mean that if the sensor perturbs the system and heavily affects the repeatability, it has an impact on the acoustic output too.

The obtained values indicate a good measurement repeatability: the average spectral standard deviations are all within the 10% of the peaks in the figure 2.19 left.
This is a very good result, giving the fact that three different measurements done on the same human subject are no comparable in frequency at all. It is nearly impossible to reproduce the same exact vocalization twice, at the same exact frequency, even for a trained singer.

From the table 2.2 emerges also that the TS configuration which exhibits the lowest values of average spectral standard deviation is the open-end configuration. That is the configuration for which the measurement exhibits higher repeatability.

2.3.4 Sensors characterization

Further experiments were carried out with the aim of underline the responses of different sensors. All the four previously introduced contact sensors were tested on the simulator, to obtain information about the frequency response and the load effect.

Frequency response

All the four transducers were tested on the TS, with the frequency sweep as source signal; the obtained spectra are presented in figure 2.20. The comparison reveals the differences in the sensors responses and confirms the necessity of a standard to be used to obtain a proper characterization. These spectra, the results obtained from this particular test, can be useful to obtain the frequency response of a contact sensor in this particular framework.

This is the effective proof that the vibration signal acquired at the base is influenced by the sensor with which the acquisition is performed. The effects are not due to the electronic or the architecture of the sensor, but it depends on geometry and dimension, which also influence the attachment method. This means that a bigger sensor, which could be more uncomfortable to wear, could have an impact on the acquired data too.
This response can be used to taking into account all the effects imputed to the contact sensor itself from the signal acquired during a vocal monitoring session, to exclude the sensor influence on the measurement. A possible extension of this work could deal about tests in which the response of the contact sensor obtained on the TS with this method is used as a filter on the vibration signal acquired on a monitoring session, performed with the Voicecare equipped with the same sensor. This could be useful to verify if the spectra presented in this section can help to increase the accuracy of the measurement performed with the Voicecare and lower the measurement uncertainty

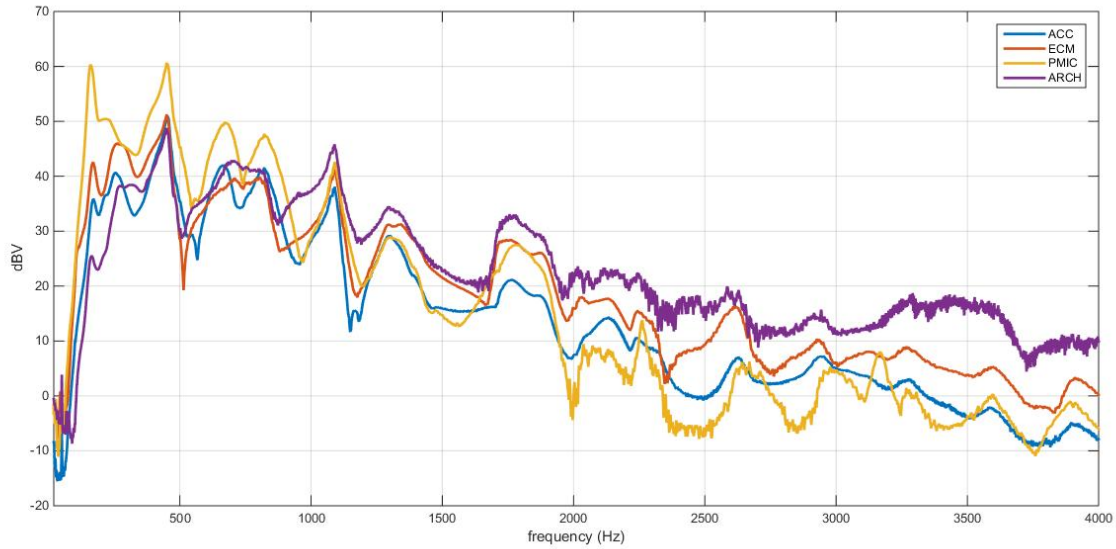


Fig. 2.20 Comparison between four different contact sensors frequency responses obtained on the TS in the open-end configuration.

estimated in [11] and [35].

Load effect estimation

As stated in the description of the TS, the embedded piezofilm has the function to sense the vibration of the phantoms without any contact sensor. This is useful to compare the vibration of the TMM of the unloaded simulator with the vibration obtained with a sensor attached on the sensing zone.
In this way, the test described in this section is meant to estimate the load effect of the four tested sensors.

Firstly, the TS without any sensor attached was driven by the frequency sweep and the piezofilm signal was acquired in order to obtain a reference spectra, i.e. the response of the unloaded simulator in each configuration. The measurement was repeated three times, the FFTs were calculated and the average spectra among the three repetitions was computed: this “unloaded” spectra can be used as a reference to estimate the load effect. The spectral standard deviation average was about 1 dBV.

The unloaded spectra were then compared with the ones obtained from the piezofilm for the repeatability study (with the TS driven by the frequency sweep sine wave). Likewise the validation test, the comparison was made by computing the differences among the points of the unloaded spectra and the loaded spectra. This can be called "spectral load effect", because it enlightens the influence of the attached sensor in the frequency domain. The average of the differences absolute values was then calculated.

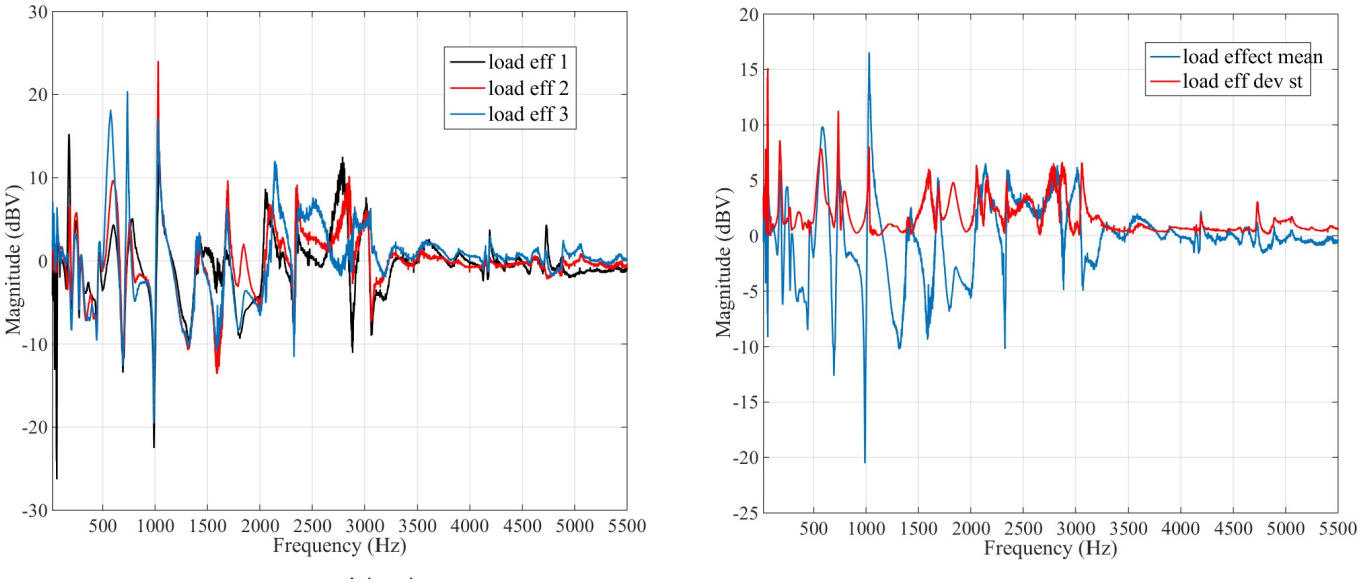


Fig. 2.21 Measurement repetition for spectral load effect evaluation (left), average spectral load effect and spectral standard deviation among the three repetitions (right). Measurements done with the ECM on the TS in open-end configuration.

As emerged from the in vivo test described in the chapter 1 and from the in vivo acquisition for the effectiveness evaluation (figure 2.16), the vibration on the neck skin originated by vocalization has its energy content in the 50÷3000 Hz interval and the frequency envelope has a peculiar shape (rectangular trapezium). Given that, the differences average among the spectra were weighted by a function which has value 1 from 50÷600 Hz, and linearly decreases from 1 to 0 in the 601÷3000 Hz interval. In this way the load effect can be estimated with a “single number”, which gives simple information about the influence of a particular sensor on the system under measurement.

Table 2.3 LOAD EFFECT VALUES AND AVERAGE SPECTRAL STANDARD DEVIATIONS FOR EVERY TESTED SENSOR IN EVERY SIMULATOR CONFIGURATION. THE CONTACT SENSOR USED FOR EVERY MEASUREMENT SERIES IS REPORTED IN THE FIRST COLUMN.

	simulator configuration	load effect (dBV)	average standard deviation (dBV)
ACC	open	0.52	2.0
	closed	0.51	2.6
	stopped	0.49	3.0
ARCH	open	0.72	2.1
	closed	0.63	3.0
	stopped	0.57	3.0
ECM	open	0.65	1.9
	closed	0.66	2.3
	stopped	0.58	2.4
PMIC	open	0.63	1.6
	closed	0.58	2.1
	stopped	0.61	2.9

An example of frequency load effect is reported in figure 2.21, and the load effect single number values are reported in table 2.3. Since every test with the sensor was repeated three times, it was possible to compute a standard deviation for every frequency load effect point. Then, the average standard deviation among the spectra was calculated. The values are reported in table 2.3.
As already seen in section 2.3.3, the TS configuration which exhibits the higher repeatability is the open-end configuration: the average standard deviation between the three repetition of the measurement reported in table 2.3 are lower for that particular TS configuration.

The ACC, which is the smaller sensor, presents the lower values of load effect. The ECM and the PMIC have similar values because of the similar shape and size. The ARCH presents the highest values because it is attached to the throat by means of a flexible plastic arch, which highly modifies the tension of the tissues near the jugular notch.

2.3.5 Conclusions

The tests carried out show that the TS mimics the phonatory system behaviour in an efficient way for the initial purpose of the work: testing contact sensors to be used in a voice monitoring framework.

The low values of the average differences between the in vivo spectra and the TS spectra presented in table 2.1 confirm the effectiveness of the approach: the vocal tract resonator amplifies the high harmonics of the source signal and creates the vowel formants, and the source signal is transmitted to the phantom surface in a similar way as a real phonatory system.

The repeatability of the measurements performed on the TS is assured by the low values of the spectral standard deviation presented in section 2.3.3: repeated measurements of the same event after the displacement and replacement of the sensor give very similar results. The simulator can be used as a standard to test and characterize contact sensors for vocal monitoring. The open-end configuration is the one which exhibits the greatest repeatability.

The embedded piezofilm stripe allows to estimate the influence of the sensor on the measurements, and the weighted average above the spectra could help to quantify this influence with a single value. The accelerometer and the piezofilm contact microphone are the sensors which exhibit the lowest load effect.
The frequency characterization points out the differences between different sensors response. This outcome and the load effect estimation results are the evidence that different sensors have a different response that must be considered in the vocal parameter evaluation. The outcame of the load effect estimation should be cosidered in the vocal monitoring results; methods to include the sensor load effect in vocal parameters estimation are still being developed.

The simulator could be useful to perform a sort of frequency calibration like the one performed on a vibrating table described in chapter 1, with the advantage that the TS is more simple to use than the vibrant table, and it could be more easily transported. As stated in section 2.2.2, the availability of a large data set of EGG, vibration at the base of the neck and voice signal simultaneously recorded, and acquired by various

and different subjects, could lead to the definition of a more accurate source signal. Moreover, it would also be useful to analyze the impact of different neck anatomies.

The TS has already been used to test the Voicecare in order to improve the calibration of the contact sensor, and to estimate the uncertainty related to the definition of the calibration function of the Voicecare device, which is also called "model error".
The model error has been obtained using the TS as a source: the reference microphone senses the voice signal at the output of the vocal tract resonator, while the contact sensor is attached to the sensing zone. The simulator has been driven by the EGG signal recorded during in vivo acquisition of a vowel /a/ at increasing intensity (part of the same acquisition was used in the effectiveness evaluation section), thus replicating the calibration procedure defined for the device.

The model error has been calculated as the maximum value, over fifteen calibrations, of the root mean square of the difference between the SPL acquired by the air microphone and the SPL estimated from the contact sensors signal.

Appendix A

References

A.1 Preface references

1 Carullo, A., Vallan, A., and Astolfi, A. Design Issues for a Portable Vocal Analyzer, IEEE Transactions on Image Processing, 62 (5), 1084-1093, (2013).

2 Carullo, A., Vallan, A. and Astolfi, A. A Low-Cost Platform for Voice Monitoring, Proceedings of I2MTC Confer-ence, Minneapolis, MN (USA), May 6-9, (2013).

3 Carullo, A., Vallan, A., Astolfi, A., Pavese, L. and Puglisi, G.E. Validation of Calibration Procedures and Uncertainty Estimation of Contact-Microphone Based Vocal, Measurement, (under review)

4 P. Bottalico, I. Ipsaro Passione, A. Astolfi, A. Carullo, E.J. Hunter, Accuracy of the quantities measured by four vocal dosimeters and its uncertainty, J. Acoustic Soc. Am., 143(3), 1591-1602 (2018)

5 A. Carullo, A. Vallan, A. Astolfi, L. Pavese, G.E. Puglisi, "Validation of Calibration Procedures and Uncertainty Estimation of Contact-Microphone Based Vocal", Measurement, vol. 74, pp. 130-142, Oct. 2015.

6 Carullo, A., Penna, A., Vallan, A., Astolfi, A., Pavese, L., Puglisi, G., Traceability and uncertainty of vocal parameters estimated through a contact microphone. IEEE MeMeA 2014 - IEEE International Symposium on Medical Measurements and Applications, Proceedings, 1-6, (2014)

7 Giordano, C., Nadalin, J., Raimondo L., Astolfi, A., Bottalico, P., Riva, G., Garzaro, M. and Pecorari G. Valutazione clinico strumentale della voce degli insegnanti ai fini della diagnosi precoce e della prevenzione delle patologie vocali, Proceedings of the 40th AIA Conference, Merano, Italy, 18-21 March, (2013).

8 Bottalico, P. and Astolfi, A. Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms, Journal of the Acoustical Society of America , 131 (4), 2817- 2827, (2012).

9 Astolfi, A., Bottalico, P., Accornero, A., Garzaro, M., Nadalin, J. and Giordano, C. Relationship between vocal doses and voice disorders on primary school teachers, Proc. Euronoise Conference, Prague, Czech Republic, 55-60, (2012).
10 Casassa F., Castellana A., Puglisi G.E., Confronto tra sensori a contatto per il monitoraggio vocale, 42 Convegno Nazioale Associazione Italiana di Acustica Proceedings (2015)

11 Castellana A., Carullo A., Casassa F., Astolfi A., Pavese L., Puglisi G.E., PERFORMANCE COMPARISON OF DIFFERENT CONTACT MICROPHONES USED FOR VOICE MONITORING ICSV 2015 Conference Proceedings (2015)

12 Casassa F., Troia A., Astolfi A., Carullo A., Vallan A., Schiavi A., Coronoa D., A phonatory system simulator for testing purposes of voice-monitoring contact sensors, Proceedings of the IEEE Internationa Instrumentation and Measurement Tecnology Conference: 1-6 (2017)

13 Casassa F., Troia A., Schiavi A., Development of a test system for voice monitoring contact sensor: phonatory system simulator, Proceedings of the International Congress on Sound and Vibration: 1-8 (2017)

A.2 First chapter references

1 Giordano, C., Nadalin, J., Raimondo L., Astolfi, A., Bottalico, P., Riva, G., Garzaro, M. and Pecorari G. Valutazione clinico strumentale della voce degli insegnanti ai fini della diagnosi precoce e della prevenzione delle patologie vocali, Proceedings of the 40th AIA Conference, Merano, Italy, 18-21 March, (2013).

2 Hunter, J. and Titze, I. R. Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings, Journal of Speech Language and Hearing Research. 53, 862–875, (2010).

3 Lyberg-Ahlander, V., Rydell, R. and Lofqvist, A. Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff, Journal of Voice, 25 (4), 430-440, (2010).

4 Bottalico, P. and Astolfi, A. Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms, Journal of the Acoustical Society of America , 131 (4), 2817- 2827, (2012).

5 Kob, G. Behler, A. Kamprolf, O. Goldschmidt, and C. Neuschaefer-Rube Experimental investigations of the influ-ence of room acoustics on the teachers voice, Acoust. Sci. and Tech., 29, 86–94, (2008).

6 Popolo P.S., Švec J.C. e Titze I.R. Adaptation of a Pocket PC for Use as a Wearable Voice Dosimeter, Journal of Speech Language and Hearing Research, 48, 780-791, (2005).

7 Cheyne, H.A., Hanson, H.M., Genereux, R.P., Stevens, K.N., Hillman, R.E. Development and Testing of a Portable Vocal Accumulator, Journal of Speech Language and Hearing Research, 46, 1457-1467, (2003).

8 Hillman, R.E. and Mehta, D.D. Ambulatory Monitoring of Daily Voice Use, ASHA Journals Perspectives on Voice and Voice Disorders, 21, 56-61, (2011).

9 Astolfi, A., Bottalico, P., Accornero, A., Garzaro, M., Nadalin, J. and Giordano, C.

Relationship between vocal doses and voice disorders on primary school teachers, Proc. Euronoise Conference, Prague, Czech Republic, 55-60, (2012).

10 Ghassemi, M., Van Stan, J.H., Mehta, D.D., Zanartu, M., Cheyne, H.A., Hillman, R.E. and Guttag, J.V. Learning to Detect Vocal Hyperfunction From Ambulatory Neck-Surface Acceleration Features: Initial Results for Vocal Fold Nodules, IEEE Transactions on Biomedical Engineering, 61 (6), 1668-1675, (2014).

11 Popolo, P.S., Rogge, M.K., Svec, J.G. and Titze, I.R. Technical considerations in the design of a wearable voice do-simeter [online] available: http://www.ncsv.org.

12 Popolo, P.S., Svec, J.G., and Titze, I.R. Adaptation of a Pocket PC for Use as a Wearable Voice Dosimeter, Journal of Speech Language and Hearing Research, 48, 780-791, (2005).

13 Svec, J.G., Titze, I.R. and Popolo, P.S. Estimation of sound pressure levels of voiced speech from skin vibration of the neck, Journal of the Acoustical Society of America, 117 (3), 1386-1394, (2005).

14 VoxLog portable voice meter [online] available: http://www.sonvox.com/index.html

15 Wirebrand, M. Real-time monitoring of voice characteristics using accelerometer and microphone measurements.

16 Cheyne, H.A., Hanson, H.M., Genereux, R.P., Stevens, K.N. and Hillman, R.E. Development and Testing of a Port-able Vocal Accumulator, Journal of Speech Language and Hearing Research, 46, 1457-1467, (2003).

17 Hillman, R.E., Heaton, J.T., Masaki, A., Zeitels, S.M., Cheyne, H.A. Ambulatory Monitoring of Disordered Voices, Annals of Otology, Rhinology and Laryngology, 115 (11), 795-801, (2006).

18 KayPENTAX Ambulatory Phonation Monitor (APM), Model 3200.

19 Carullo, A., Vallan, A., and Astolfi, A. Design Issues for a Portable Vocal

Analyzer, IEEE Transactions on Image Processing, 62 (5), 1084-1093, (2013).

20 Carullo, A., Vallan, A. and Astolfi, A. A Low-Cost Platform for Voice Monitoring, Proceedings of I2MTC Confer-ence, Minneapolis, MN (USA), May 6-9, (2013).

21 Švec Palacký, J. G., Granqvist, S. Guidelines for Selecting Microphones for Human Voice Production Research, American Journal of Speech-Language Pathology, 19, 356–368 , (2010).

22 Stevens, K., Kalikow, D., and Willemain, T. A miniature accelerometer for detecting glottal waveforms and nasali-zation, Journal of Speech Language and Hearing Research, 18, 594-599, (1975).

23 Lippmann, R. P., Detecting nasalization using a low-cost miniature accelerometer, J. S. Hear. Res, 24, 314–317, (1981).

24 Carullo, A., Vallan, A., Astolfi, A., Pavese, L. and Puglisi, G.E. Validation of Calibration Procedures and Uncertainty Estimation of Contact-Microphone Based Vocal, Measurement, (under review).

25 Zanartu, M., Ho, J.C., Kraman, S.S., Pasterkamp, H., Huber, J.E. and Woducka, G.R. Air-Borne and Tissue-Borne Sensitivities of Bioacoustic Sensors Used on the Skin Surface, IEEE Tr. on Bi.En., 56 (2), 443- 451, (2009).

26 Hillman, R.E., and Metha, D.D. Ambulatory monitoring of Daily Voice Use, Perspective on Voice and Voice Disorders, 21(2), 56-61, (2011).

27 P. Bottalico, I. Ipsaro Passione, A. Astolfi, A. Carullo, E.J. Hunter, Accuracy of the quantities measured by four vocal dosimeters and its uncertainty, J. Acoustic Soc. Am., 143(3), 1591-1602 (2018)

28 Hillman R.E., Heaton T.J., Masaki A., Zeitels S.M., Cheyne H.A., Ambulatory monitoring of disordered voices. Annals of Otology, Rhinology and Laryngology, 115(11), 795-801 (2006)

29 Hillman, R.E., Metha D.D., Ambulatory monitoring of daily voice use, Perspective on voice and voice disorders, 21(2), 56-61 (2011)

A.3 Second chapter references

1 D.D. Metha, M. Zañartu, S. W. Feng, H. A. Cheyne, R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform", IEEE Trans. On Biomedical Engineering, vol. 59, n. 11, pp. 3090-3096, Nov. 2012

2 G. J. Švec, I. R. Titze, P. S. Popolo, "Estimation of sound pressure levels of voiced speech from skin acceleration of the neck", J. Acoust. Soc. Am, vol. 117, no. 3 pt. 1, pp 1386-1394, March 2005

3 M. Ghassemi, J. H. Van Stand, D. D. Metha, M. Zañartu, H. A. Cheyne, R. E. Hillman, J. V. Guttag, "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: initial results for vocal fold nodules", IEEE Trans. on biomedical engineergin, vol. 61, no. 6, June 2014, pp. 1668-1675

4 P. Bottalico, A. Astolfi, "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms", Journal of the Acoustical Society of America, vol. 131 no. 4, pp. 2817- 2827, 2012.

5 A. Astolfi, p: Bottalico,A. Accornero, M. Garzaro, J. Nadalin, C. Giordano, "Relationship between vocal doses and voice disorders on primary school teachers", Proc. Euronoise Conference, Prague, Czech Republic, pp. 55-60, 2012.

6 M. Wirebrand, "Real-time monitoring of voice characteristics using accelerometer and microphone measurements"

7 H.A. Cheyne, H.M. Hanson, R.P. Genereux, K.N. Stevens, R.E. Hillman, "Development and Testing of a Portable Vocal Accumulator", Journal of Speech Language and Hearing Research, vol. 46, pp. 1457-1467, 2003.

8 P.S. Popolo, J.C. Švec, I.R. Titze, “Adaptation of a Pocket PC for Use as a Wearable Voice Dosimeter”, Journal of Speech Language and Hearing Research, vol. 48, pp. 780-791, 2005.

9 A. Carullo, A., A. Vallan, A. Astolfi “Design Issues for a Portable Vocal Analyzer”, IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 5, pp 1084-1093, 2013.

10 A. Carullo, A. Vallan, A. Astolfi, “A Low-Cost Platform for Voice Monitoring”, Proceedings of I2MTC Conference, Minneapolis, MN (USA), May 6-9, 2013

11 A. Carullo, A. Vallan, A. Astolfi, L. Pavese, G.E. Puglisi, “Validation of Calibration Procedures and Uncertainty Estimation of Contact-Microphone Based Vocal”, Measurement, vol. 74, pp. 130-142, Oct. 2015.

12 M. Zañartu, C.C. Ho, S. S. Kraman., H. Pasterkamp, J. E. Huber, G.R. Wodicka, “Air-Borne and Tissue-Borne Sensitivities of Bioacoustic Sensors Used on the Skin Surface”, IEEE Tr. on Biomedical engineergin, vol. 56, no. 2, pp.443- 451, 2009.

13 Y.D. Heman-Ackah, D.D. Michael, G.S. Goding, “The relationship between cepstral peak prominence and selected parameters of dysphonia”, J Voice,vol.16, pp. 20–27, 2002.

14 Y. Maryn, M. De Bodt, N. Roy, “The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders,” J. Commun. Disord.,vol. 43, pp.161–174, 2010.

15 S.N. Awan, N. Roy, M.E. Jette, G.S. Meltzner, R.E. Hillman, “Quantifying dysphonia severity using a sepctral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V,” Clin. Linguist. Phon., vol.24, pp.742–758, 2010.

16 A. Carullo, F. Casassa, A. Castellana, A. Astolfi, L. Pavese, G. E. Puglisi, “Performance comparison for different contact-microphones used for voice mointoring”,

22nd Interntion congress on sound and vibration, Florence (ITA), June 12-16, 2015.

17 S.S. Kraman, G. A. Pessler, H. Pasterkamp, G. R. Wodicka, “Desing, construction and evaluation of bioacoustic transducer testing (BATT) system for respiratory sound”, IEEE Trans. on biomedical engineering, vol. 53, no. 8, pp. 1711-1715, Aug. 2015.

18 S.S. Kraman, G. R. Wodicka, G. A. Pessler, H. Pasterkamp, “Comparson of lung sound transucers using a bioacoustic transducers testing system”, J. Appl. Physiol., vol. 101, pp. 469-476, Aug. 2006.

19 S. F. Austin, I. R. Titze, “The effect of subglottal resonances upon vocal fold vibration”, Journal of voice, vol. 11, iss. 4, pp. 391-402, December 1997.

20 J. Horacek, V. Uruba, V. Radolf, J. Vesely, V. Bula, “Airflow visualization in a model of human glottis near the self-oscillating vocal folds model”, J. Apllied and computational mechanics, vol. 5, pp 21-28, 2011.

21 L. P. Fulcher, R. C. Sherer, T. Powell, “Pressure distribuition in a static physical model of the uniform glottis: entrance and exit coefficients”, J. Acoust. Soc. Am., vol. 129, no. 3, March 2001.

22 G. C. J. Hofmans, G. Groot, M. Ranucci, G. Graziani, A. Hirshberg, “unsteady flow through in-vitro model of the glotts”, J. coust. Soc. Am, vol. 113, no. 3, march 2003.

23 T. Vampola, J. Horacek, J. G. Švec, “FE modeling of human vocal tract acoustics. Part 1: production of czech vowels”, Acta acustica united with Acustica, vol. 94, pp 433-447, 2008.

24 A. Schiavi, R. Cuccaro, A. Troia, “Strain-rate and temperature dependent material properties of Agar and Gellan Gum used in biomedical applications”, Journal of the mechanical behavior of biomedical materials, vol. 53, p. 119-130, 2015.

25 R. Cuccaro, C. Musacchio, PA. Giuliano Albo, A. Troia, S. Lago, “Acoustical

characterization of polysaccharide polymers tissue-mimicking materials", Ultrasonics, vol. 56, p. 210-219, 2014.

26 K.J.M. Surry, H.J.B. Austin, A. Fenster, T.M. Peters, "Poly(vinyl-alcohol) cryogel phantoms for use in ultrasound and MR imaging", Physics in Medicine and Biology, vol. 49, no. 24, pp. 5529-5546, 2004.

27 Zañartu, M., Ho, JC., Mehta DD., Hillman, R.E., Wodick,a G.R., Subglottal Impedance-Based Inverse Filtering of Voiced Sounds Using Neck Surface Acceleration, IEEE Trans Audio Speech Lang Process, 21(9), 1929-1939 (2013)

28 Harper, P., Kraman, S. S., Pasterkamp, H., Wodicka, G. R., An Acoustic Model of the Respiratory Tract, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, 48 (5), 543-550 (2001)

29 Carullo, A., Penna, A., Vallan, A., Astolfi, A., Pavese, L., Puglisi, G., Traceability and uncertainty of vocal parameters estimated through a contact microphone. IEEE MeMeA 2014 - IEEE International Symposium on Medical Measurements and Applications, Proceedings, 1-6, (2014)

30 Chen, A.I., Balter, M.L., Chen, M.I., Gross, D., Alam, S.K., Maguire, T.J., Yarmush, M.L., Multilayered tissue mimicking skin and vessel phantoms with tunable mechanical, optical, and acoustic properties, Med Phys, 43(6), 3117-3131 (2016)

31 Culjat M.O., Goldenberg D., Tewari P., Singh R.S., A review of tissue substitutes for ultrasound imaging, Ultrasound Med Biol, 36(6), 861-873 (2010)

32 Maccabi A., Taylor Z., Bajwa N., Mallen-St Clair J., St John M., Sung S., Grundfest .1, Saddik G. An examination of the elastic properties of tissue-mimicking phantoms using vibro-acoustography and a muscle motor system, Rev Sci Instrum, 87(2) (2016)

33 Cook J.R., Bouchard R.R., Emelianov1, S.Y., Tissue-mimicking phantoms for photoacoustic and ultrasonic imaging, Biomed Opt Express, 2(11), 3193–3206 (2011)

34 Humphreys B.K., Delahaye M., Peterson C.K., An investigation into the validity of cervical spine motion palpation using subjects with congenital block vertebrae as a 'gold standard', BMC Musculoskelet Disord. 5, 19 (2004)

35 P. Bottalico, I. Ipsaro Passione, A. Astolfi, A. Carullo, E.J. Hunter, Accuracy of the quantities measured by four vocal dosimeters and its uncertainty, J. Acoustic Soc. Am., 143(3), 1591-1602 (2018)

36 Casassa F., Troia A., Astolfi A., Carullo A., Vallan A., Schiavi A., Coronoa D., A phonatory system simulator for testing purposes of voice-monitoring contact sensors, Proceedings of the IEEE Internationa Instrumentation and Measurement Tecnology Conference: 1-6 (2017)

37 Casassa F., Troia A., Schiavi A., Development of a test system for voice monitoring contact sensor: phonatory system simulator, Proceedings of the International Congress on Sound and Vibration: 1-8 (2017)

38 Hillman R.E., Heaton T.J., Masaki A., Zeitels S.M., Cheyne H.A., Ambulatory monitoring of disordered voices. Annals of Otology, Rhinology and Laryngology, 115(11), 795-801 (2006)

39 Hillman, R.E., Metha D.D., Ambulatory monitoring of daily voice use, Perspective on voice and voice disorders, 21(2), 56-61 (2011)

Appendix B

Schematic and graphs

B.1 Contact sensors performance comparison

The conditioning circuits used for the tests discussed on the first chapter are reported below.

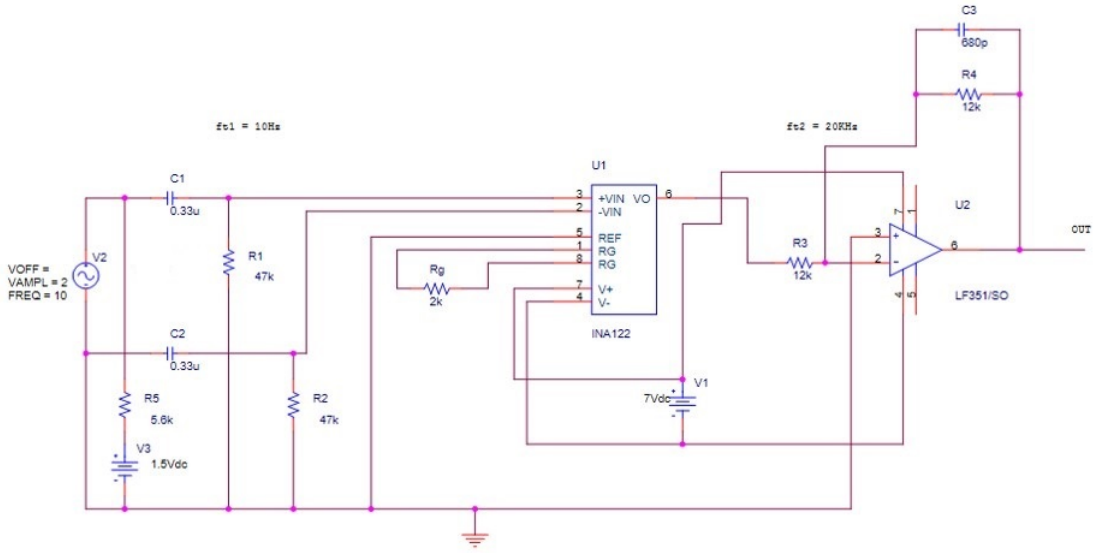


Fig. B.1 Accelerometer conditioning circuit.

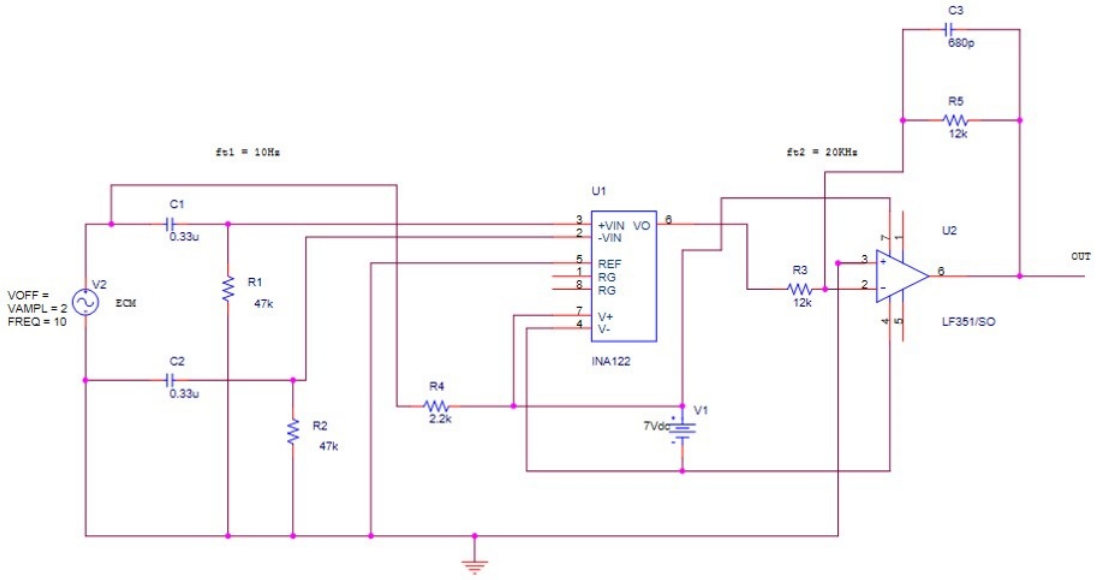


Fig. B.2 ECMs conditioning circuit.

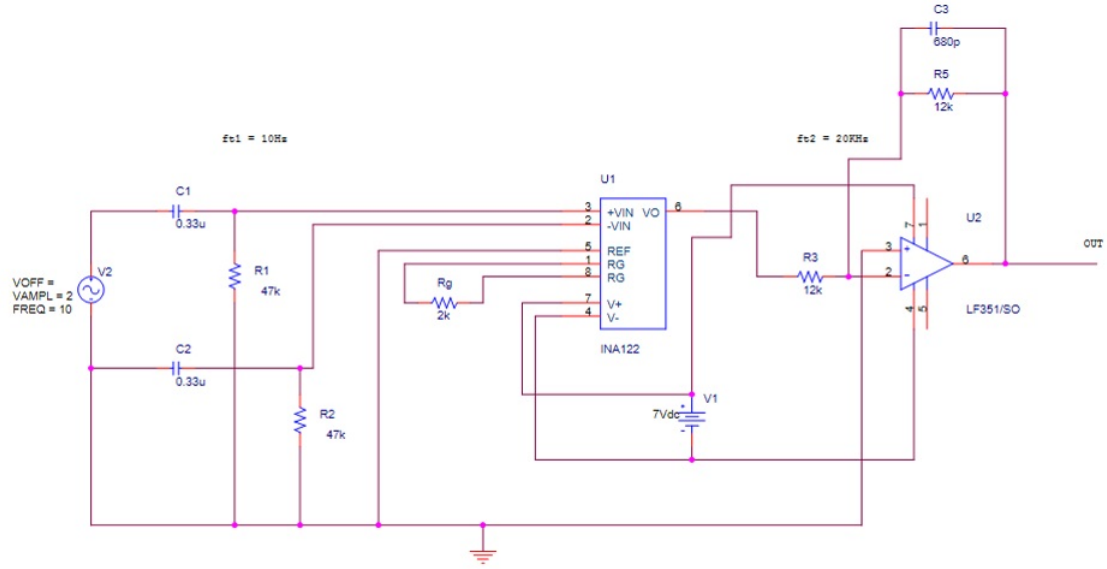


Fig. B.3 Piezoelectric transucer conditioning circuit.

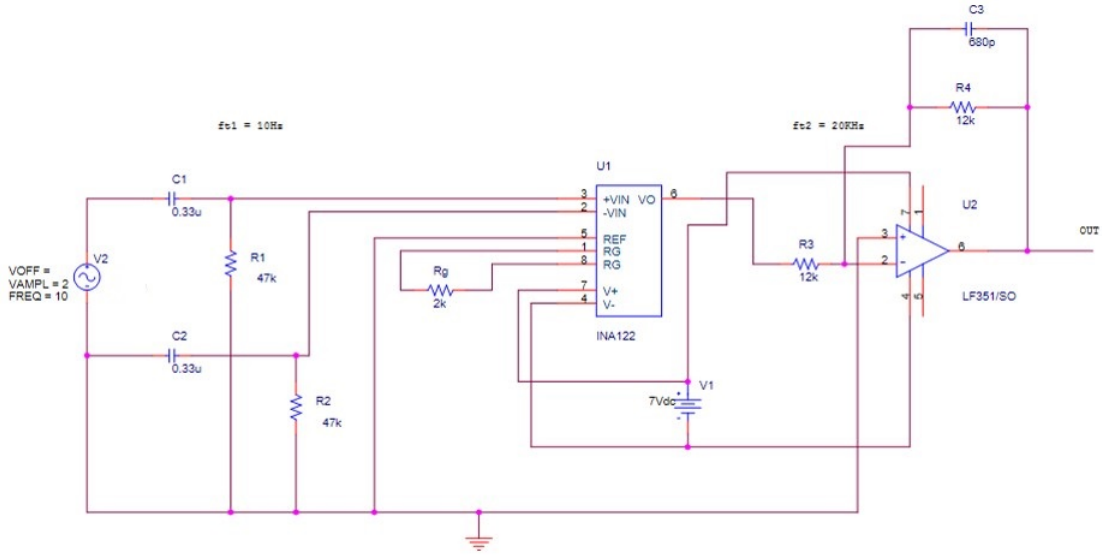


Fig. B.4 Reference microphone conditioning circuit.

B.2 TS effectiveness evaluation

In this section the graphs obtained during the effectiveness evaluation test of the TS are reported. For every contat sensor - TS configuration combination 3 graphs are presented: the comparison between the spectra acquired invivio and on the TS with the contact sensor (left), the same comparison for the air microhone spectra (center) and the difference between the two acquisition for the two channels.
For the closed-end configuration, it is noticeable how the difference between the invivo spectra and simulator spectra have their minimum near the harmonics, that is where the negative peaks occurr. This fact is more evident in figure B.6.
In almost every configuration, in the air microphone average spectral difference between the invivo spectra and the simulator spectra, the higher peaks are always in the same frequency, for every configuration and for every used sensor. This confirm what is stated in section 2.3.2: the high values in the air microphone spectral difference depends on the fact that the artificial vocal tract used on the simulator is different from the vocal tract of the subject monitored for this purpose, but the stable peaks in the average spectral differences confirm that the response of the acoustic output of the simulator is stable.

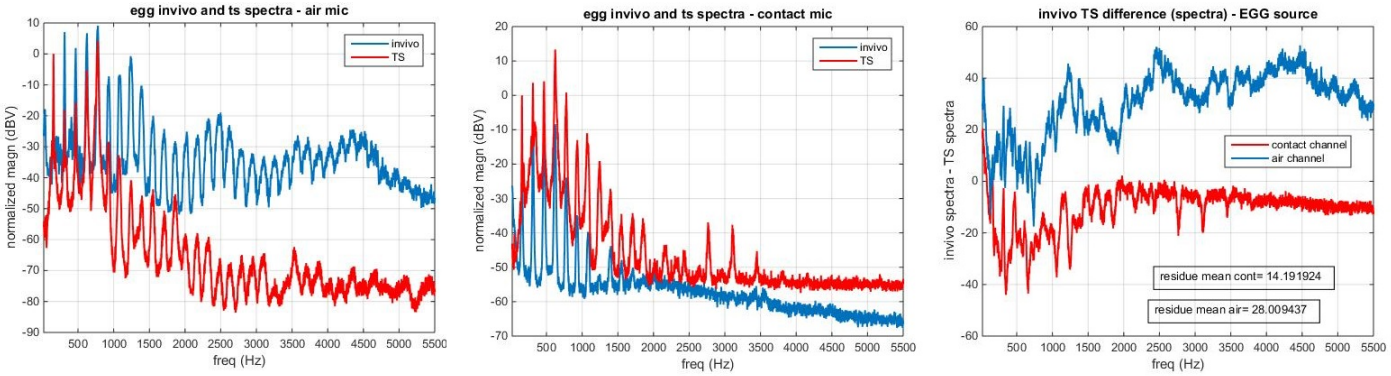


Fig. B.5 Accelerometer - closed-end configuration effectiveness evaluation.

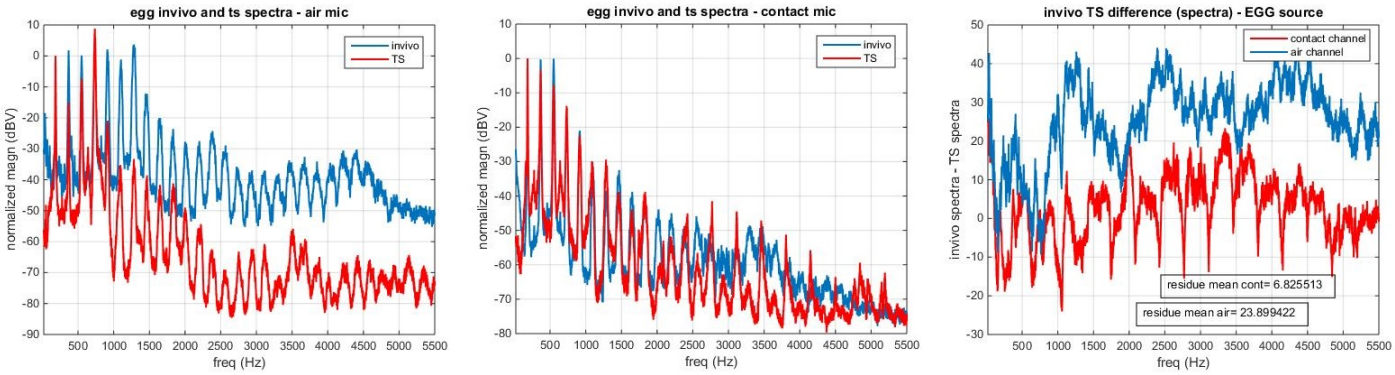


Fig. B.6 ECM - closed-end configuration effectiveness evaluation.

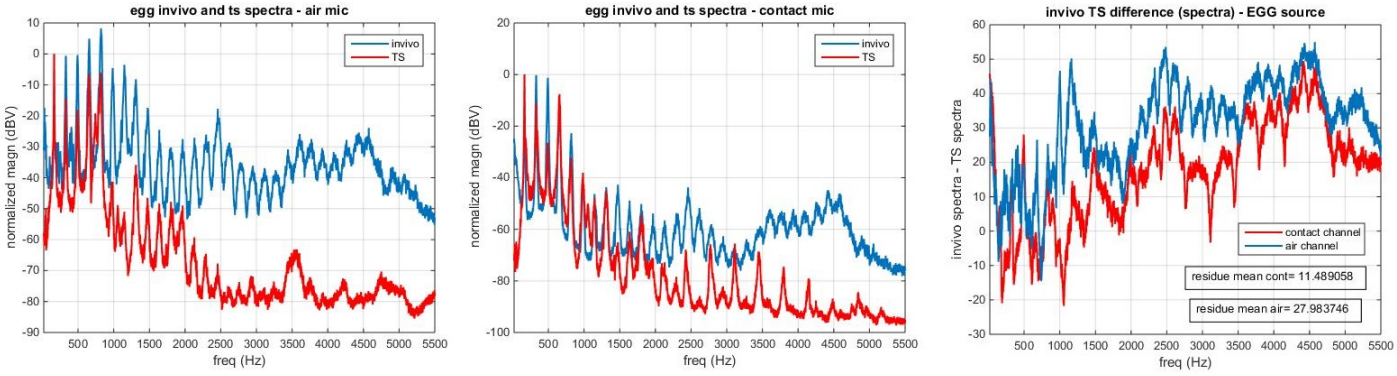


Fig. B.7 Piezofilm contact mic - closed end configuration effectiveness evaluation.

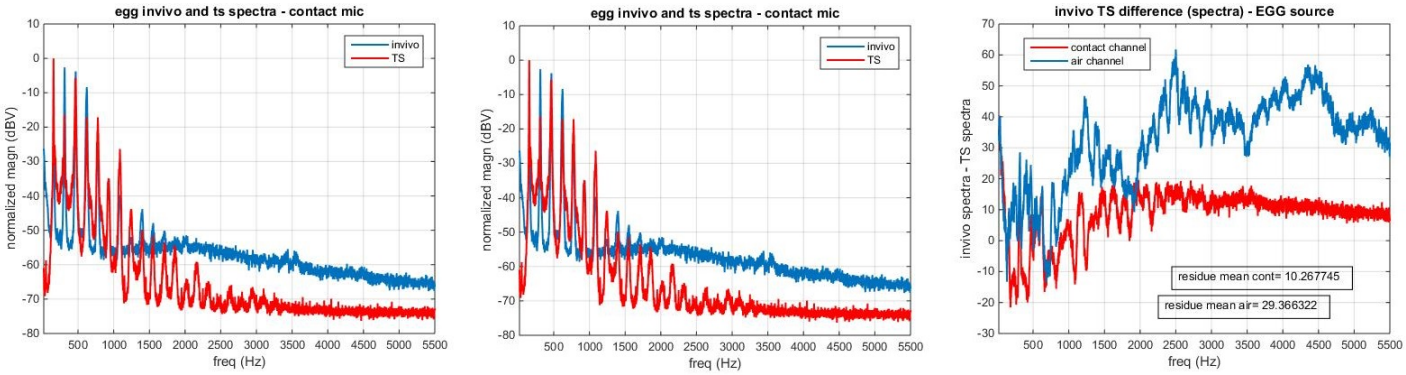


Fig. B.8 Accelerometer - open-end configuration effectiveness evaluation.

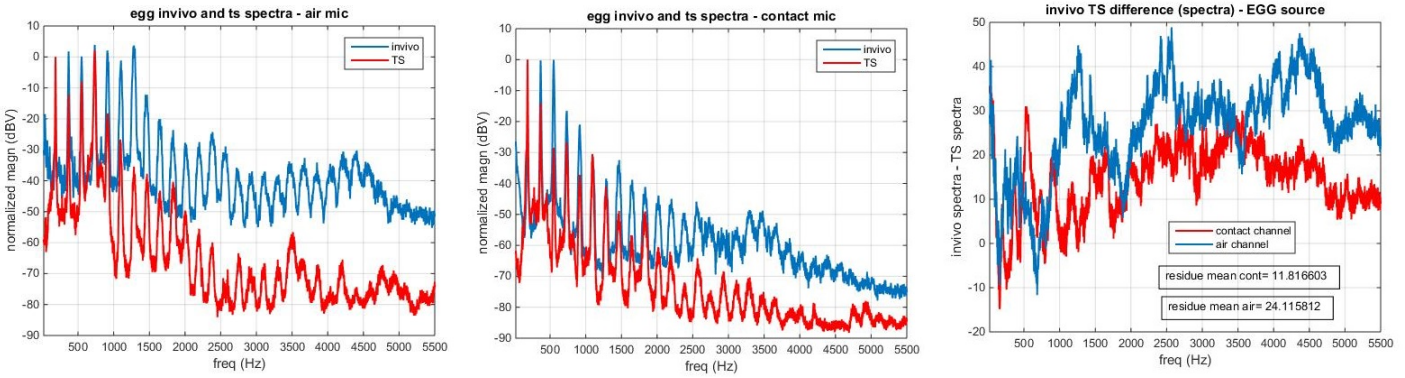


Fig. B.9 ECM - open-end configuration effectiveness evaluation.

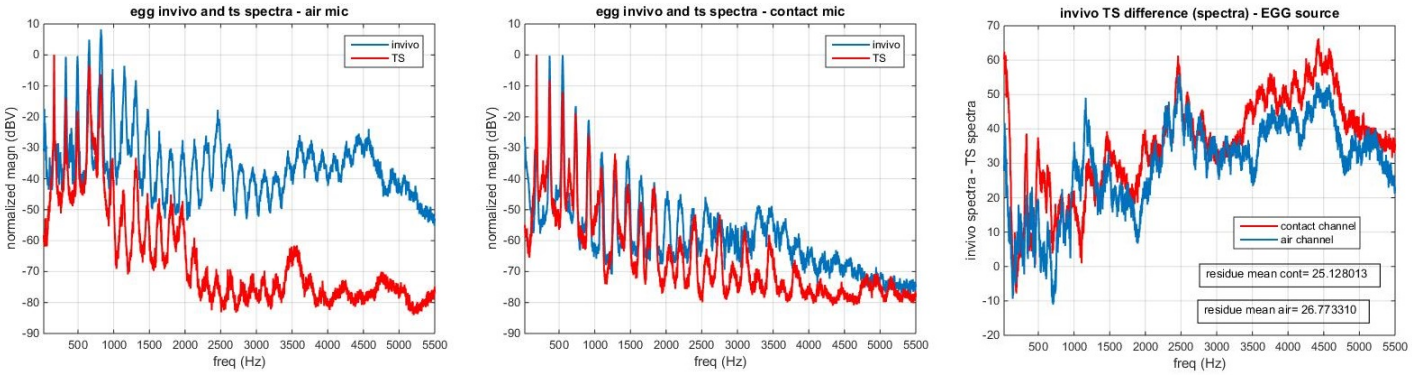


Fig. B.10 Piezofilm contact mic - open-end configuration effectiveness evaluation.

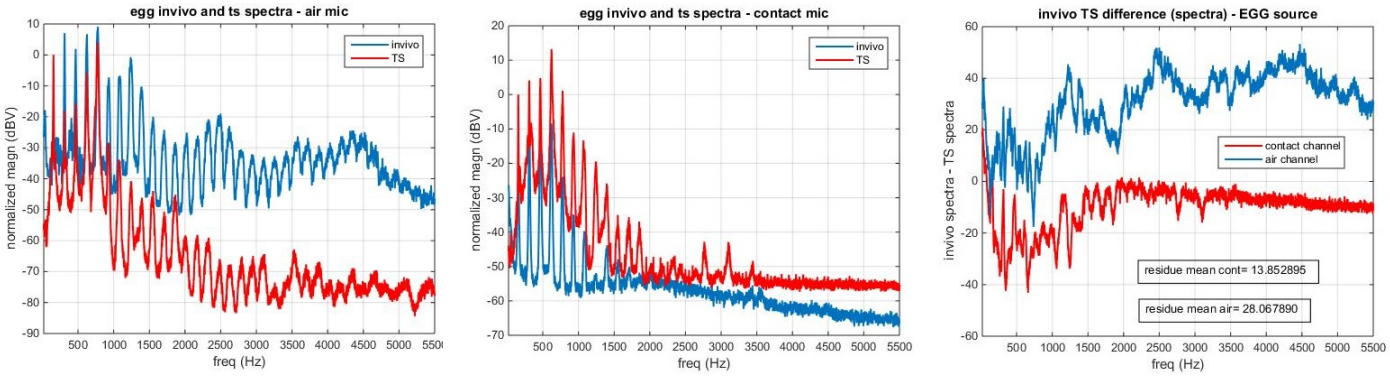


Fig. B.11 Accelerometer - stopped-end configuration effectiveness evaluation.

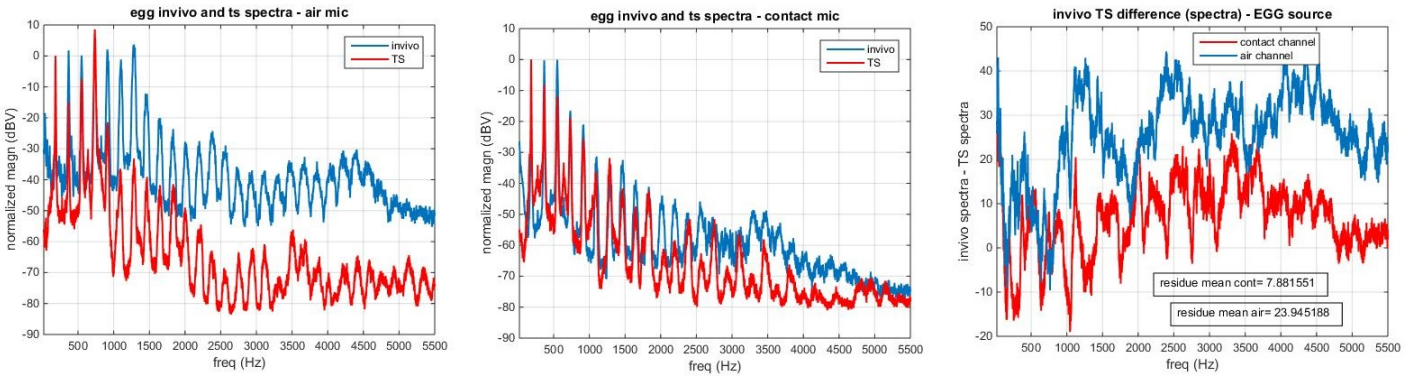


Fig. B.12 ECM - stopped-end configuration effectiveness evaluation.

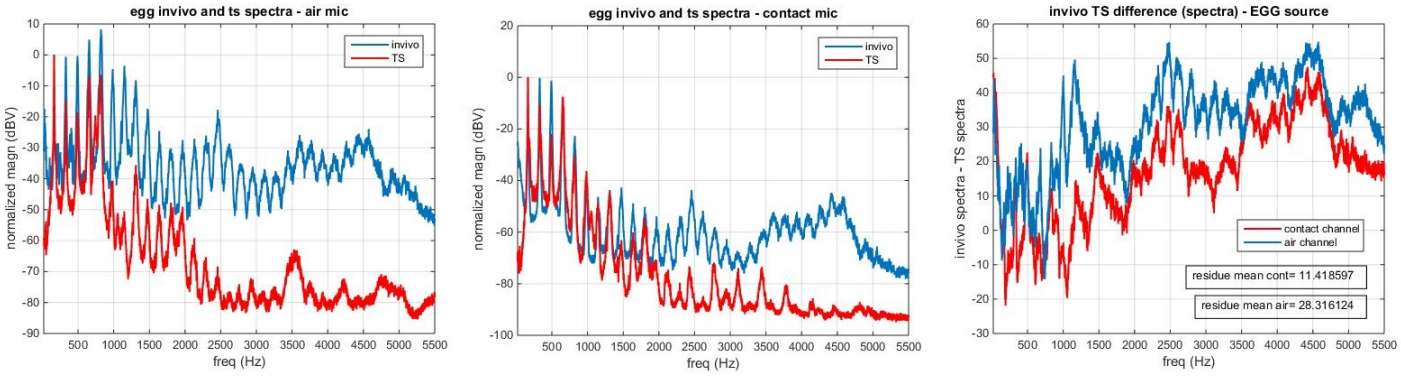


Fig. B.13 Piezofilm contact mic - stopped-end configuration effectiveness evaluation.

B.3 Sensors' load effect estimation

In this section, the graph obtained from the contact sensors load effect estimation on the TS are reported. In left graphs, the spectra of the three repetition of the load effect measurement are reported; right graphs show the average spectra and the standard deviation of the spectral points among the three repetitions.
The important data are reported in tab 3.3, the "single numnber" load effect, which is reported in every graph with average spectra and standard deviation (right graph); nevertheless, it is important to notice that the average load effect spectra present peaks (relative maximum and minimum)in corrispondence of the peaks present in the original source signal. This means that the load effect of a contact sensor, in this particular application, changes the response of the object of measurement (the tissues at the base of the neck) mainly at those frequencies that are the key to describe the acquired signal, and which are used to estimate various frequency-based parameters, like cepstral peak prominence CPP.

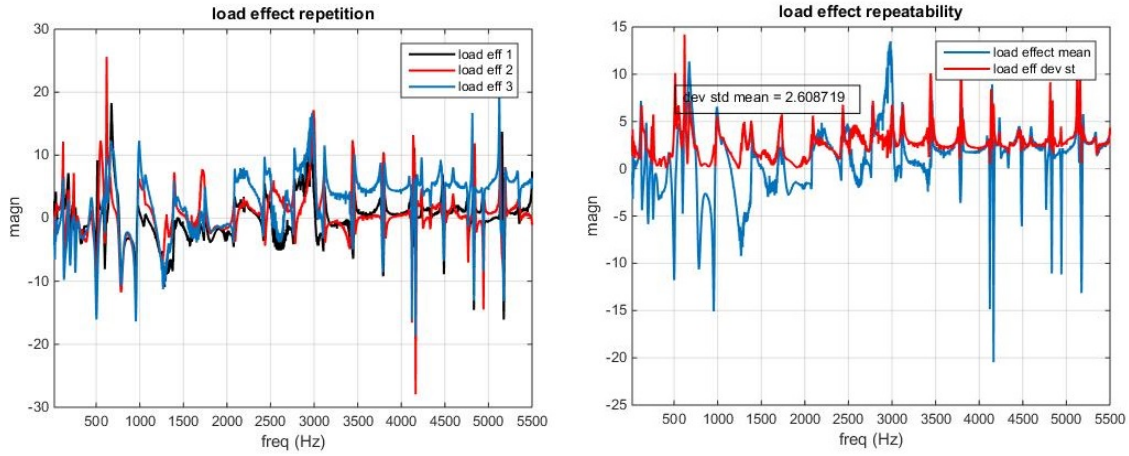


Fig. B.14 Accelerometer load effect measurement, closed-end configuration.

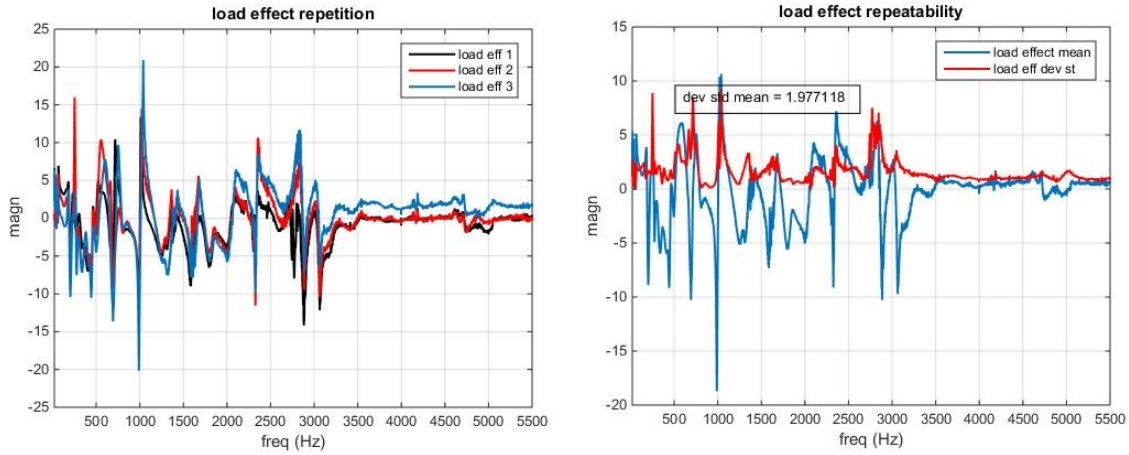


Fig. B.15 Accelerometer load effect measurement, opend end configuration.

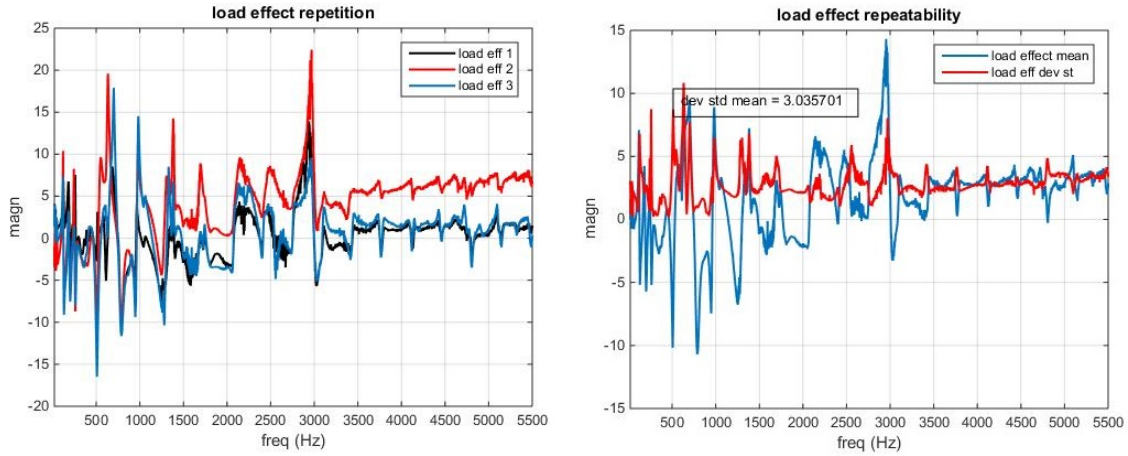


Fig. B.16 Accelerometer load effect measurement, stopped-end configuration.

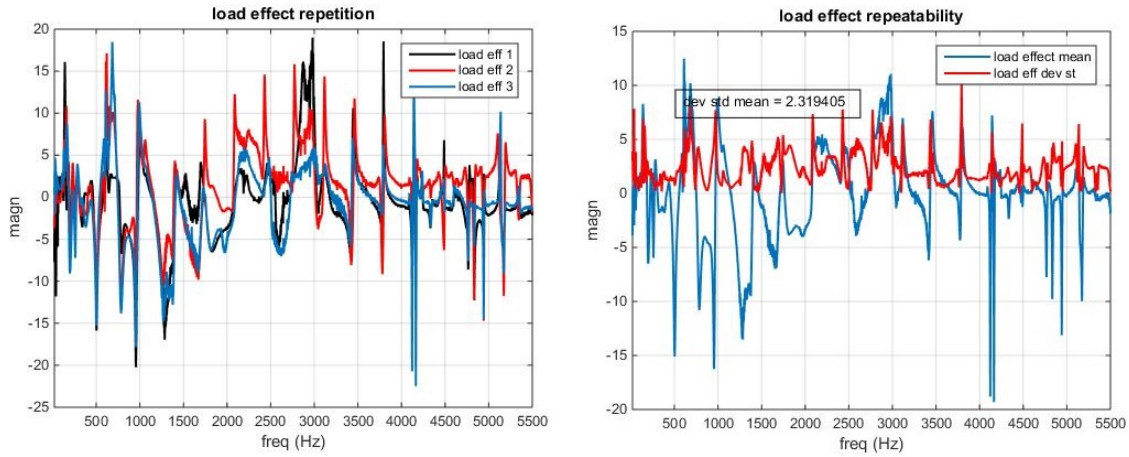


Fig. B.17 ECM load effect measurement, closed-end configuration.

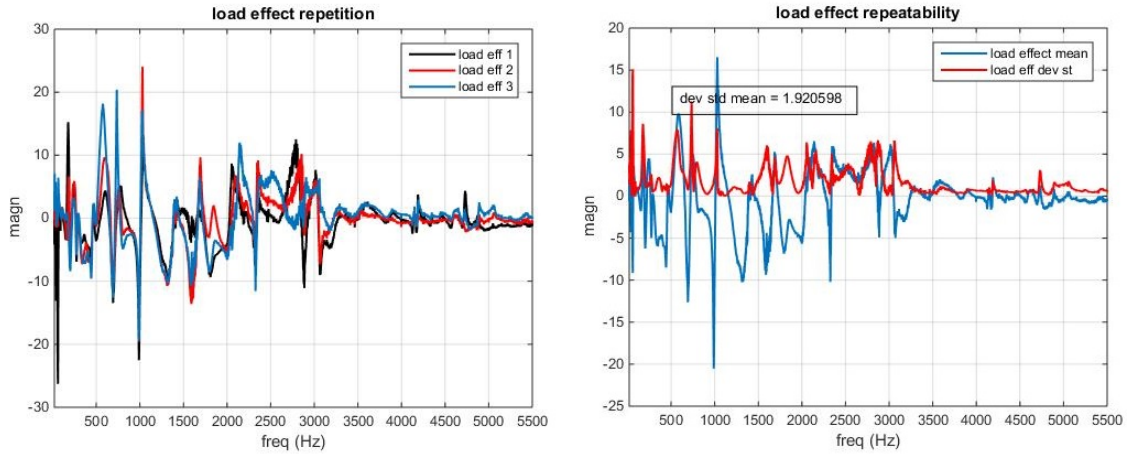


Fig. B.18 ECM load effect measurement, open-end configuration.

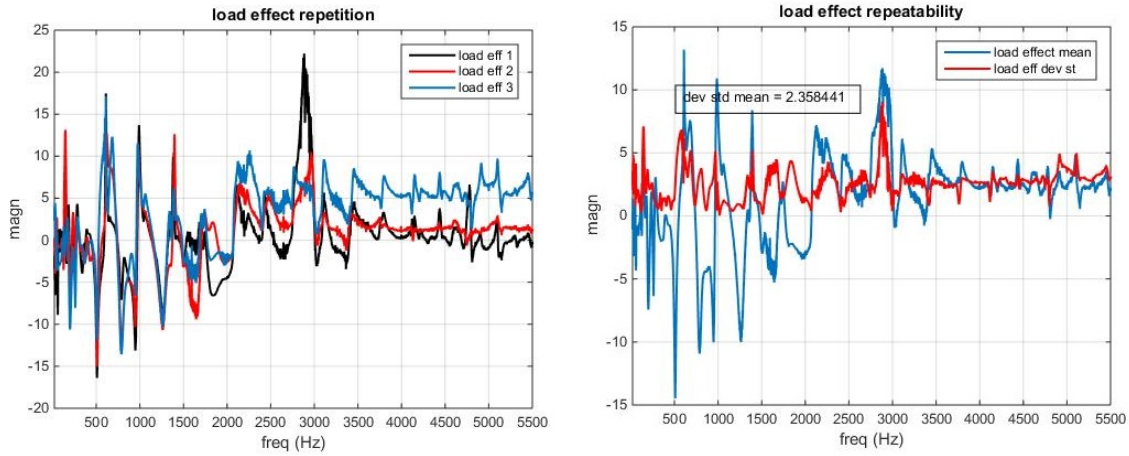


Fig. B.19 ECM load effect measurement, stopped-end configuration.

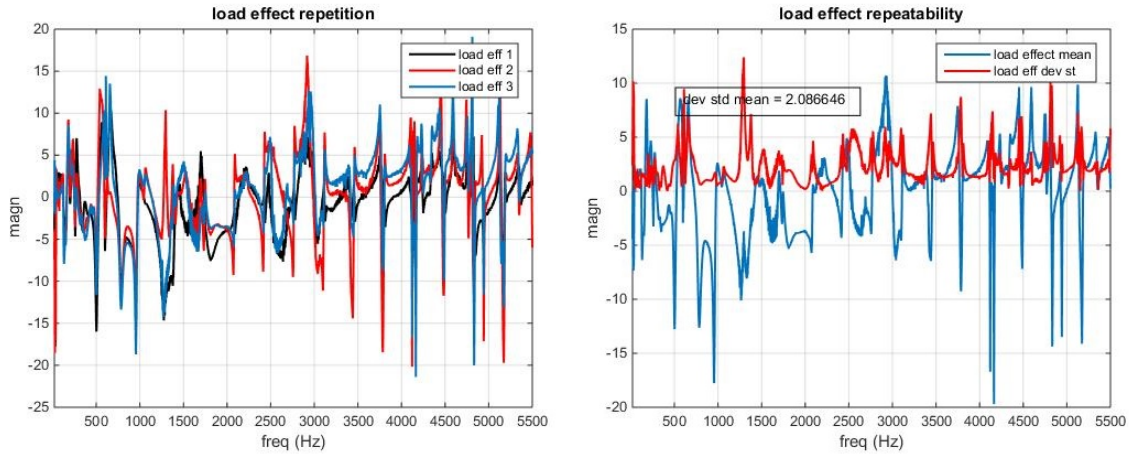


Fig. B.20 Piezofilm contact microphone load effect measurement, closed-end configuration.

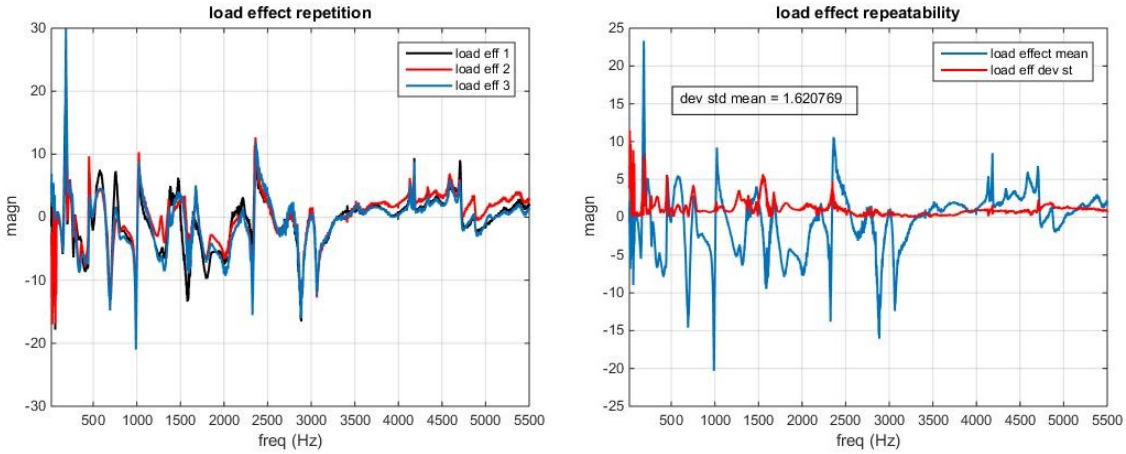


Fig. B.21 Piezofilm contact microphone load effect measurement, open-end configuration.

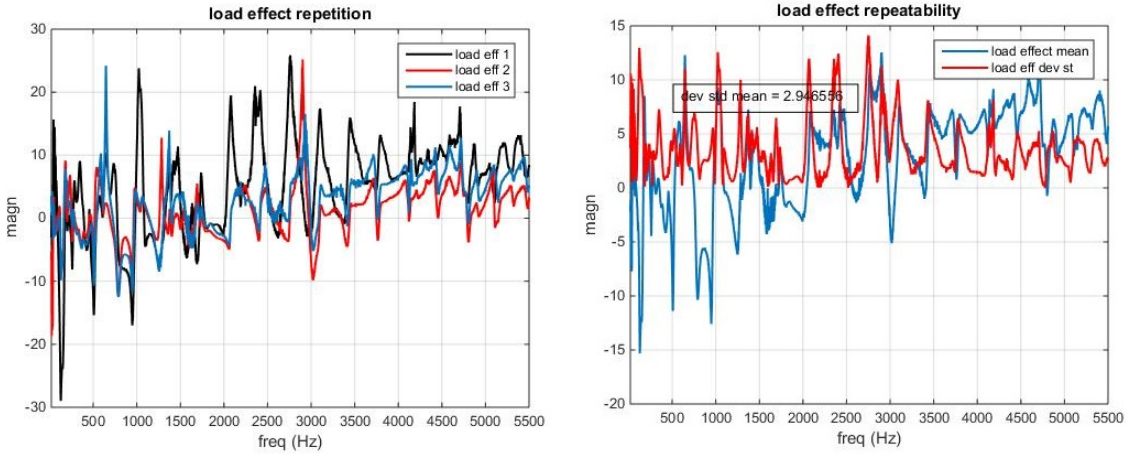


Fig. B.22 Piezofilm contact microphone load effect measurement, stopped-end configuration.

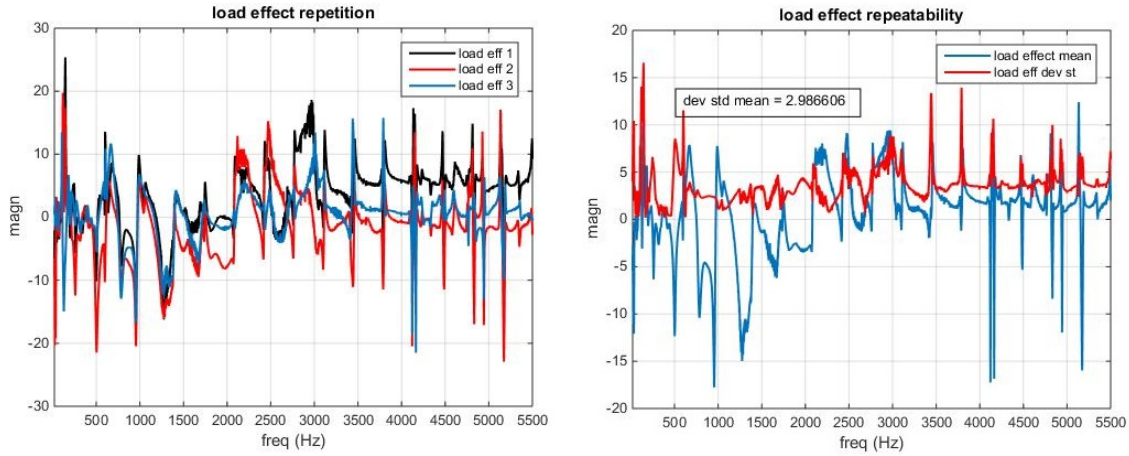


Fig. B.23 Piezoelectric throat microphone load effect measurement, closed-end configuration.

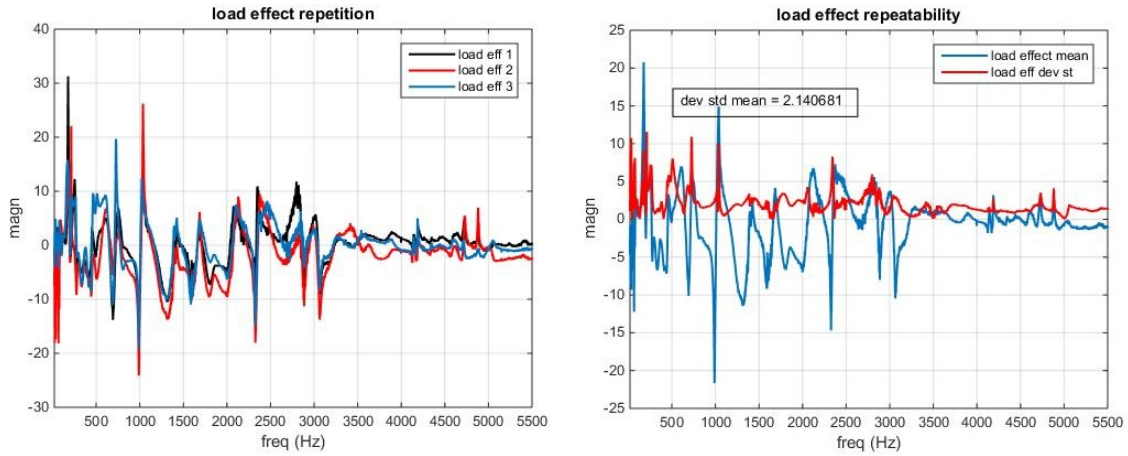


Fig. B.24 Piezoelectric throat microphone load effect measurement, open-end configuration.

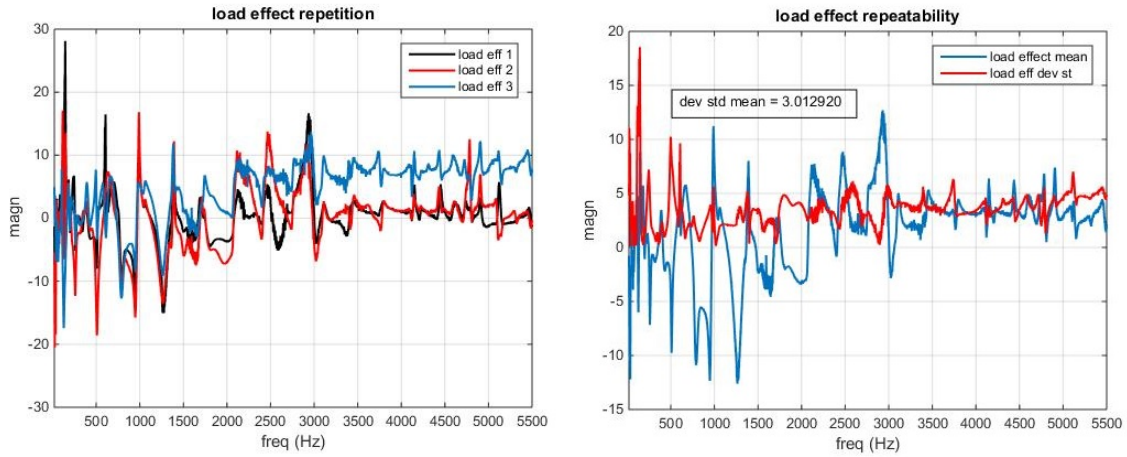


Fig. B.25 Piezoelectric throat microphone load effect measurement, stopped-end configuration.

B.4 Sensors' frequency response

In this section, the graph obtained from the contact sensors frequency response test on the TS discussed in section 3.3.4 are reported.

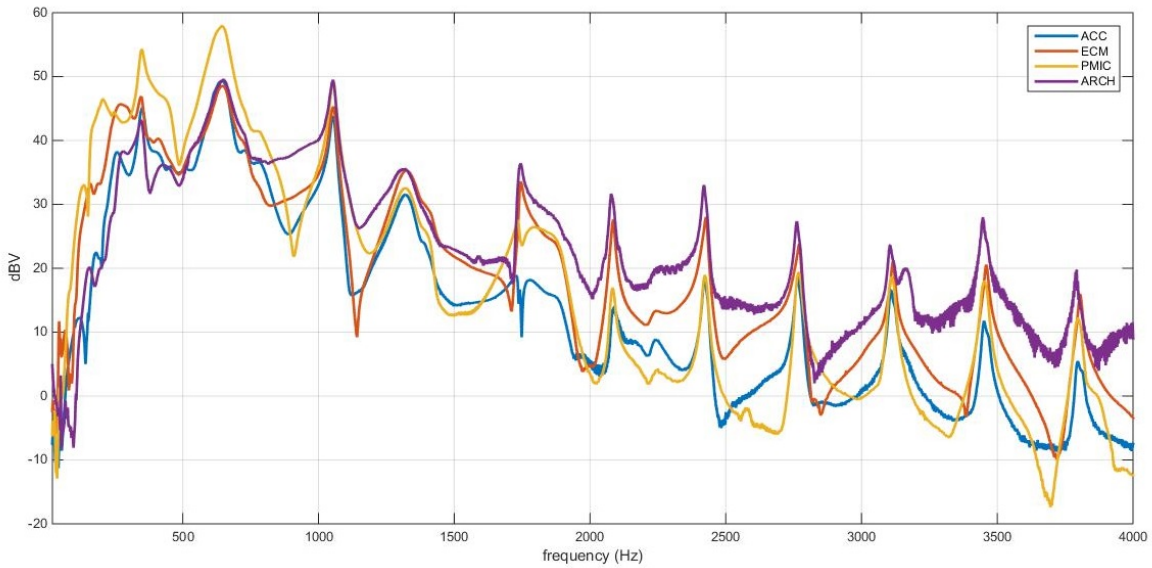


Fig. B.26 Contact sensors' frequency response, closed-end configuration.

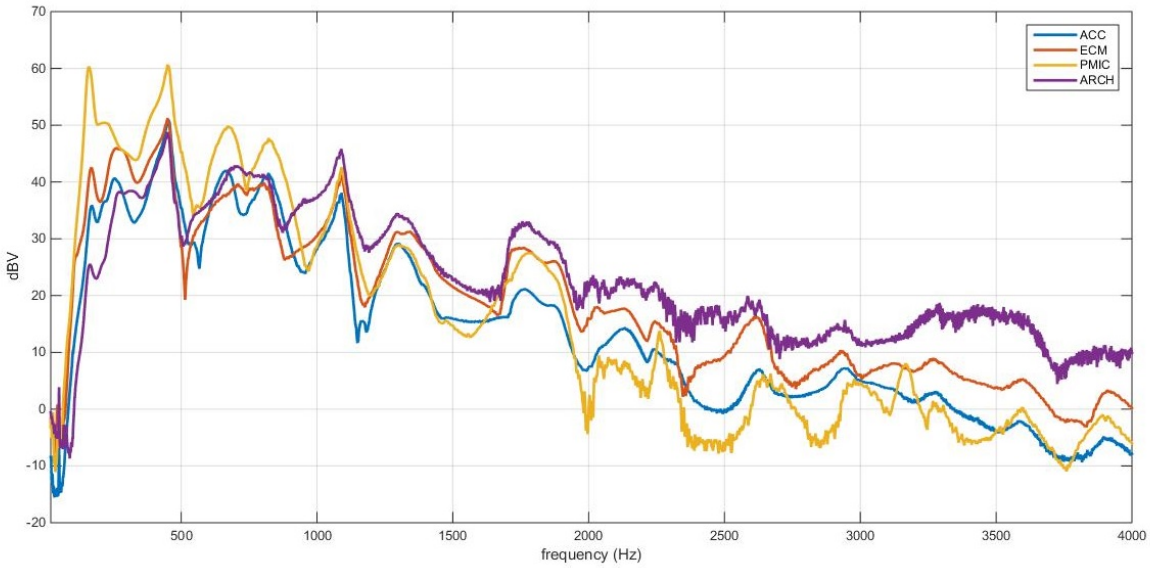


Fig. B.27 Contact sensors' frequency response, open-end configuration.

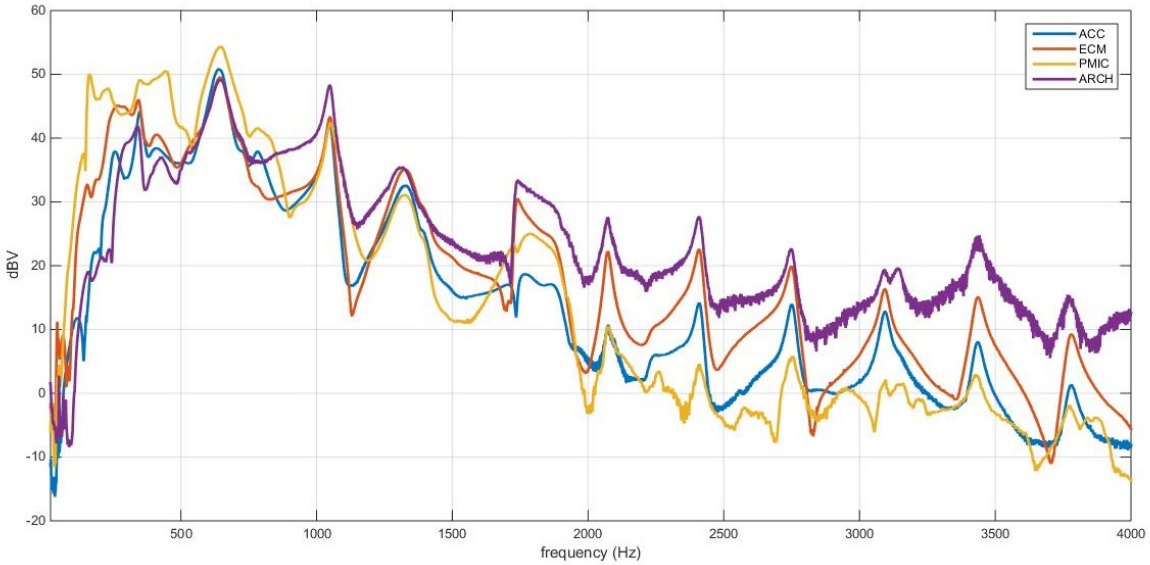


Fig. B.28 Contact sensors' frequency response, stopped-end configuration.

Appendix C

Background theory

In this appendix some fundamental definitions and concepts of acoustics will be presented and explained. This appendix is useful to understand how a sound can be produced, how it propagates and the quantities used for describe its features.
Then the phonatory system will be described, the source of the phenomena object of this research, and one of the lead actors in this dissertation. The material presented in this appendix is taken from [1].

C.1 Fundamentals of acoustics

In this first section some fundamental definitions and concepts of acoustics will be presented and explained. This introduction is essential to understand how a sound can be produced, how it propagates and the quantities used for describe its features. The material of this appendix is Generally speaking, a sound is generated whenever there is a disturbance of the equilibrium of density (or pressure) in a medium that could be a gas, liquid or solid. For what concern the human ear, the medium needed to perceive this disturbances is the air. If the pressure disturbance is positive in respect of the average pressure/density, there will be a condensation in the medium, an increase in air density. If the pressure disturbance is negative, there will be a rarefaction, a decrease in air density.

In figure 1 three different sound sources are presented. For the piston in a cylinder (fig. c.1 a) air is alternatively condensed and rarified by oscillatory movement of

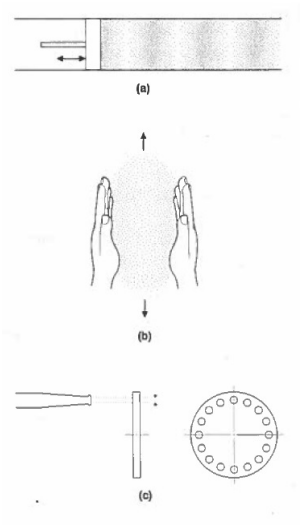


Fig. C.1 Sources of sound: piston in a cylinder (a), hand clapping (b) and siren (c)

the piston. The particles in contact with the piston attempt to follow the oscillatory motion, but since they are not rigidly connected to the piston surface, they establish their own characteristic motion by collision. Momentum is imparted to adjacent particles in a time delayed fashion, creating a new disturbance at a point further away from the piston. This "passing along" of pressure disturbances by the medium is called sound *propagation*.

Figure c.1 b illustrates handclapping. Sound is produced by the sudden interruption of the airflow that is squeezed out between the hands. This transient disturbance of airflow can be amplified by cupping the hands slightly, thereby forming a resonance chamber.

Siren is another source of sound (fig. c.1 c). A jet of air is directed toward a series of holes on the perimeter of a rotating wheel. The period interruptions of the airflow through the holes create large pressure disturbances, and hence very intenze sound. The quicker the flow is started and stopped, the higher the frequency components of the sound will be. The air is forced through a periodically opening and closing orifice, and the result is a series of flow pulses. The number of pulses in the unit of time is the frequency of the produced sound, and the amount of airflow driven by the source is related to the intensity of the produced sound.

All acoustical phenomena can be described in time as a waveform, but this is not always an efficient way to express it, especially when the sound to be described is complex. In this case the spectra (in the frequency domine) can be a more easy-to-

read way to express the characteristics of a sound. Given any sound, the spectra shows the number of component frequencies and the strength of each component. This components represent all the oscillation of the local pressure disturbance in the medium. Perceptually, a single component frequency is a pure tone, and its waveform is a single sinusoid (which can be described by three numbers: frequency, amplitude and phase). But in reality almost every sound is composed by more than one component frequency, and in general an infinite collection of sinusoids is needed to construct a complex tone.
Let's see now how the sound can be described in terms of what we called "intensity" before. The change of the static pressure can be espressed as $p_{tot} = p_{stat} + p$ where p is the deviation from the static ambient pression which represents the sound. The intensity of a sound can be calculated as $I = pv$, where v is the so called *particle velocity*, which is the speed with which the medium particle oscillates to create the perturbation.
But the descriptor that is usually used to describe the intensity of a sound is the *sound pressure level (SPL)* L_p, which is a logarithmic measure of the pressure of a certain sound related to a reference pressure level p_0:

$$L_p = 20log_{10}\left(\frac{p}{p_0}\right) dB \tag{C.1}$$

where

- $p_0 = 20\mu Pa$ is the reference pressure;
- p is the root mean sound pressure.

The unit of measure is the Decibel, which is a logarithmic unit used for levels.

C.2 The speech production

Voice is our primary means of expression and is perceived by two of the five senses. We not only hear voice, but we also feel it: the unborn children grown accustomed to the vibrations they receive in the womb when the mother vocalizes. When listening to the music, our desire to "feel" the sound is sometimes strong, partly because it

relates to the way we feel our own voice productions. Vibrations in the head, neck and chest are indications that sound is produced. Regarding the voice production:

- *voice* has both a narrow and a broad definition. In the broad sense, voice is synonymous with speech. In the narrow sense, it refers only to the sound produced by the vocal fold vibration. With this definition, all the sounds we make for speech are either voiced or unvoiced. If the vocal folds vibrate, the sound is voiced, regardless of what other simultaneous sounding may occurr (hissing, tongue flapping, clicking etc.).
- *Vocalization* also refers to sound produced by vocal fold vibration, but the term is most applicable to nonspeech or prespeech soundings, like sounds produced by animals and infants and by singers when they warm up or sing a song without words.
- *Phonation* is a technical term used to describe the physical and physiological processes of vocal fold vibration

This section is dedicated to the description of the human phonatory system and of how it works to produce the phonation.

C.2.1 Phonatory system physiology

In the figure c.2 a scheme of the larynx is presented. In the pictures only cartilages, bones, two tracheal rings and some ligaments and membranes are illustrated. Not shown in the figure are muscles, blood vessels, nerves and other soft tissue that surrounds cartilages. But a real larynx never looks like figure c.2, because it is difficult to tell where one type of tissue begins and another ends.
The laryngeal cartilages are the thyroid cartilage (whose protrusion present in males is called Adam's apple), the cricoid cartilage, the arytenoid cartilage which containes the vocal folds, and the epiglottis. The last two, which are the ones strictly involved in the phonation, will be described in details.
The two arytenoid cartilages are situated on top on wider posterior portion of the cricoid cartilage (see combination view in figure c.3 d). Near the base of each arytenoid cartilage there are two projections, or processes: posteriorly there is the muscular processes and anteriorly there is the vocal processes. The vocal process is

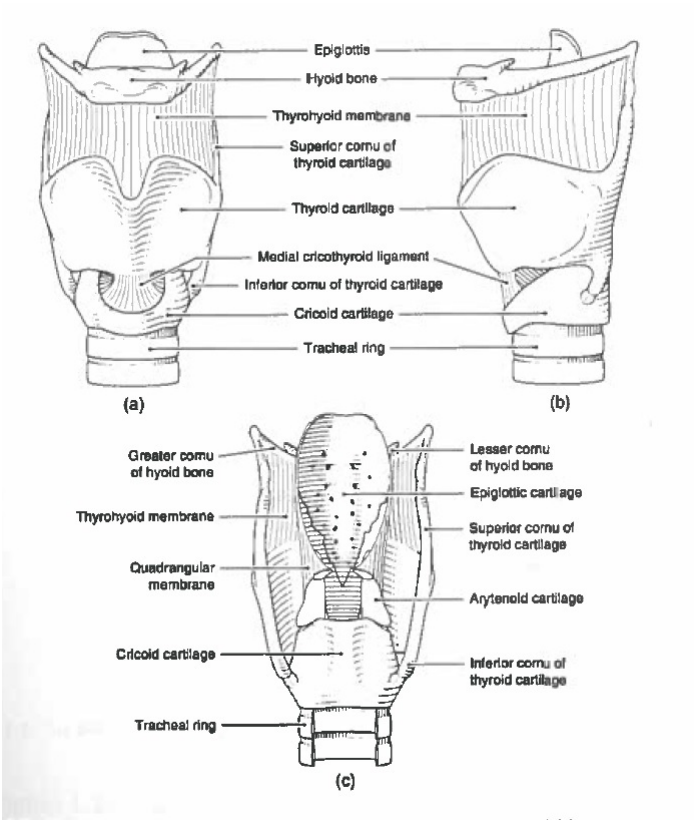


Fig. C.2 Larynx: anterior view (a), lateral view (b), posterior view (c).

the point of attachment of the vocal lingament, an important part of the vocal folds. The muscula and vocal processes can be positioned in a variety of ways to *abduct* (move apart) or *adduct* (bring toghether) the vocal folds. This action is accomplished through the highly flexible cricoarytenoid joint.

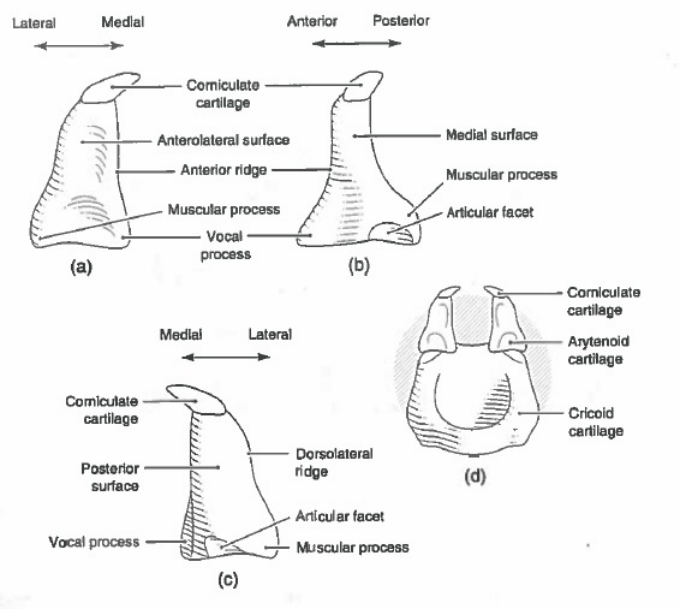


Fig. C.3 Arytenoid cartilages: anterior view (a), lateral view (b), posterior view (c) and composite anterior view with cricoid cartilage (d).

The articular facet is a curved surface to allow a rocking motion (rotation and translation) of the arytenoid cartilage on top on the cricoid. In figure c.3 d it can be seen that the attachment of the arytenoid cartilage to the cricoid cartilage is such that permits this rocking motion in the joint.
Thus, the arytenoid cartilage can move not only in the medial-lateral direction but also in the anterior-posterior direction.

The epiglottis is showed in figure c.4. This cartilage resembles the tongue of a shoe and has a somewhat similar function. As a tongue of a shoe folds over the foot when the shoe is laced, the epiglottis folds over the entryway to the larynx when tight closure of the airway is desired. By attachment whit its connective tissue to the inner surface of the tyroid cartilage, just below the tyroid notch, the epiglotis forms the anterior wall of a chamber. This epiglottal chamber collapses during food

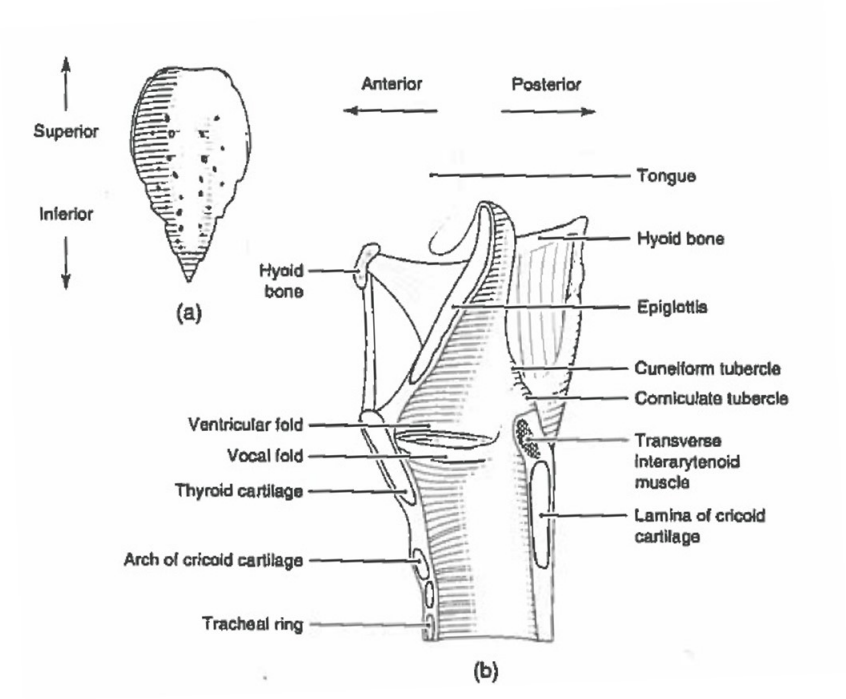


Fig. C.4 Epiglottis: posterior view (a), midsagittal section through the larynx showing attachments of the epiglottis to the hyoid bone, tongue, and thyroid cartilage (b).

transport but can serve as an acoustic resonator when the airway is open. Superiorly, the epiglottis attaches to the base of the tongue; it also attaches to the hyoid bone via the hyoepiglottic ligament.

The hyoid bone, which is not properly a part of the larynx, is a horseshoe shaped structure that can be seen in figure c.2 and c.5. The Hyoid bone partially surrounds the tip of the epiglottis. Vertically, it connects to the thyroid cartilage through the thyrohyoid membrane and the superior cornu. Many muscles are anchored to this bone, and this is the exact function of this particular bone: be a sort oh hitching point for tissue connection that are needed to form an approximate 90° bend in the airway from the mounth to the pharynx. The bone also helps to protect some of the soft tissues in the upper larynx and lower pharynx against injuries that might occur from blows to the neck.

Many of the structures of the larynx can be viewed with videolaryngoscopy: a rigid fibreoptic cable is introduced into the mouth, or a flexible fiberoptic cable is introduced into the nose in order to observe the larynx. The flexible cable allows the patient under examination to produce phonation during the examination. This tecnique allows the physicians to observe the larynx "in action", in order to identify possible strange behaviours.

Another examination method useful to observe the larynx is the laryngostroboscopy, in which a strobe light is combined with rigid or flexible laryngoscopy. This tecnique allows the observation of the motion of the vocal cords by the method of indirect laryngoscopy through the use of intermittent light. Laryngostroboscopy is conducted by means of a special instrument called the laryngostrobo-scope, which permits the frequency of the light pulses to be adjusted to the frequency of vibration of the vocal cords of the person being examined. The adjustment is automatic in the modern digital laryngostroboscopes. If the frequency of the light pulses coincides with the frequency of the vocal cords, the cords appear to be motionless. The vibrations of the vocal cords become visible when the frequency of the light pulses is made unequal to that of the vocal cords. Laryngostroboscopy is used in the determination of functional and organic lesions of the larynx.
Quoting Kosztyła-Hojna and Rogowski [2]:

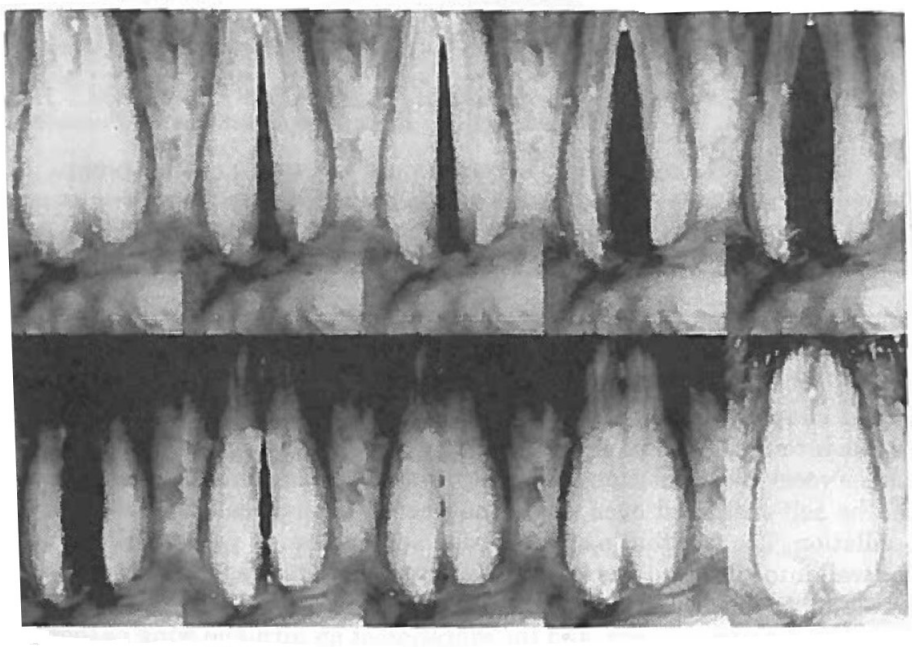


Fig. C.5 Vocal folds viewed by videolaryngoscopy: abduction (a) and adduction (b).

"The observation of the glottis and the vocal fold mobility during phonation enables the diagnosis of larynx pathology. Videolaryngostroboscopy (VLSS) facilitates acquiring a precise endoscopic picture and an evaluation of the vocal fold vibratory movements. This method is recognised as an objective, repetitive and non-invasive approach to accelerate early diagnosis in laryngeal carcinoma, vocal nodules, vocal fold paresis, larynx oedema, functional dysphonia and presbyphonia. [...] Aberrations in the vocal fold vibrations indicate a supraepithelial oedema of the laryngeal mucosa and a functional type of dysphonia, requiring differential therapy. The larynx image recorded on a video tape is a valuable diagnostic evidence that allows monitoring of therapeutic effects and phoniatric rehabilitation."

The muscles of the larynx (figure c.6) can be divided in two groups: intrisic and extrinsic. The intrinsic muscles interconnect the cartilages of the larynx, and are the ones more involved in the phonation process.
The thyroarytenoid muscle links the thyroid cartilage to the arytenoid cartilage, and it makes up the bulk of the vocal folds. The result of contraction of this muscle is to draw the arytenoid cartilage forward and so shortening and thickening the vocal folds. The crycotyroid muscle is the so called pitch-control muscles: it works in pair

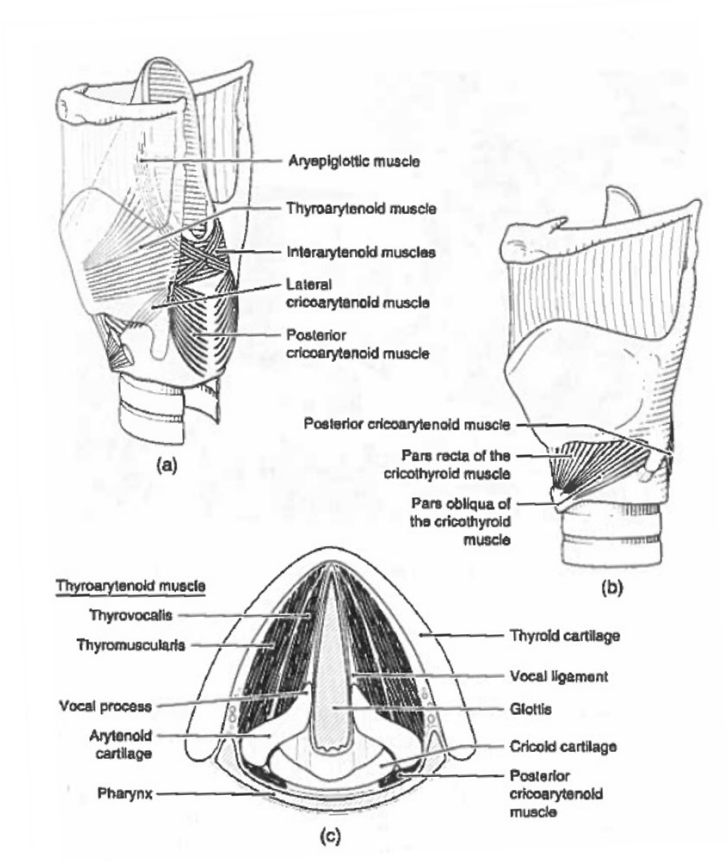


Fig. C.6 Intrinsic muscles of the larynx: posterior-lateral view (a), anterior-lateral view (b) and superior view (c).

with the thyroarytenoid muscle and it is responsible of the lenghtening of the vocal folds.
The lateral cryctoarytenoid contraction joins toghether the vocal folds in the front. It works in pair with the posterior cryctoarytenoid muscle, which is the primary abductor of the vocal folds.
The interarytenoid muscle connects the two arytenoid cartilages, and it is considered to have two parts. This muscle serves as adductor of the vocal folds.

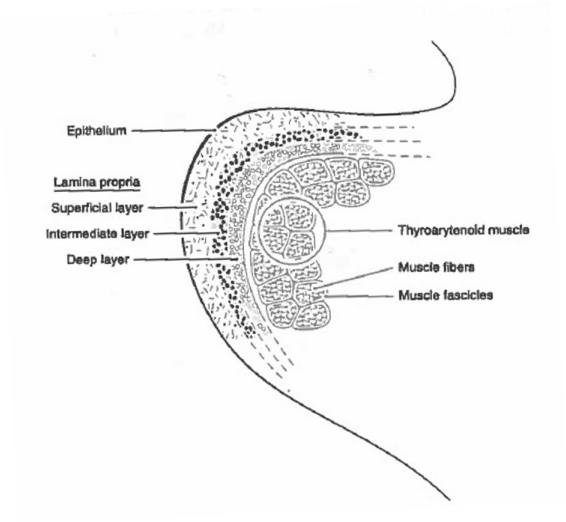


Fig. C.7 Schematic of the coronal section of the right vocal fold, showing soft tissue layers.

The vocal fods are located at the narrowest portium of the airway between the trachea and the tip of epiglottis (see figure c.4). In combination with the ventricular folds, the aryepiglottic folds and the quadangular membrane constitute a system of folds that seals off the laryngeal airway rapidly and completely when the appropriate muscles are activated. Figure c.7 shows a drawing of the layered structure of the right vocal fold in coronal section. The outermost layer is made up of stratified squamous epithelium, which encapsulates softer, fluidlike tissue, somewhat like a balloon filled with water.
The lamina propria, a layered system of nonmuscular tissues, is between the epithelium and muscles. It can conveniently divided into three layers: superficial (loosely organised elastin fibers surrounded by interstitial fluids), intermediate (elastin fibers uniformely oriented in the longitudinal direction) and deep (collagen fibers, nearly

inextensible, which limit elongation).

C.2.2 Phonation mechanics

The mechanical working of the phonatory system is based on the movement of the vocal folds: a self-sustained oscillation, or *flow-induced oscillation*.
Classical description of the vocal folds vibration usually begins with the assertion that the vocal folds are sucked together by a negative pressure in the glottis. According to the flow energy conservation, this action is possible if the glottis is sufficiently narrow, the ariflow is sufficiently high and the glottal wall is soft enough to yeld. The collapse of the glottis is then followed by a buildup of the subglottal pressure during closure, causing the vocal folds to begin to move laterally and the glottis to open. Lateral movement continues until elastic forces in the tissue retard the motion and ultimately reverse it. The tissue then moves medially again, and another cycle of collapse begins.

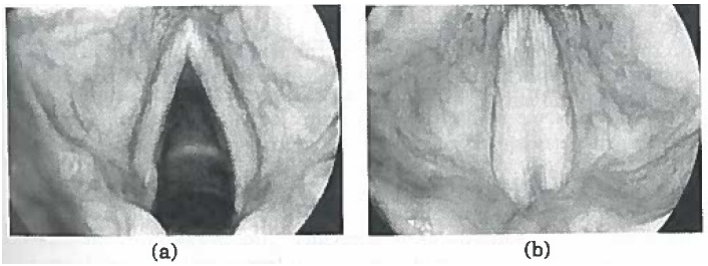


Fig. C.8 A series of video frame of vocal fold movement during a normal glottal cycle.

This movement is showed in figure c.8, in which are presented a series of frame taken with a laryngostroboscope. The description of vocal folds vibration invoking the Bernoulli effect (negative pressure in the glottis), tissue elasticity and vocal folds collision has been called the *myoelastic-aerodynamic theory of the vocal fold vibration*. This theory is however inadeguate in explaining the important features of self-sustained oscillation. The mechanism for continual energy transfer from the airstream to the tissue involves more than the Bernoulli forces alone, which not distinguish between inward and outward movement of the vocal folds. A special vibration mode of the tissue, or a coupling to the vocal tract, is needed to allow the

Bernoulli forces to lower during glottal closing and to raise during glottal opening. Otherwise, not enough energy will be imparted to the tissue and oscillation will damp out. The subglottal pressure applied to the vocal folds during glottal closure can also serve as a driving force, but this would have nothing to do with Bernoulli forces.

Two mechanical systems frequently used to study oscillation are the pendulum and a mass attached to a spring. A child on a playground swing is an interesting example of pendulum, and can help to understand the concept of self-sustained oscillation. This concept can be applied to the mass-spring oscillator, which is a better mechanical model of the vocal folds.
A self sustained mechanichal oscillator has to meet three conditions: there must be an equilibrium position, there must be inertia in the system to overshoot the equilibrium position and the net energy loss per cycle of oscillation must be zero. Consider the child on a swing again: after an initial boost, the wind resistance, dragging feet and the friction in the joints between the swing frame and the chains are responsible for energy loss, and the inertia of the system is not enough to compensate this loss. If the child "pumps" energy in the system (by moving the torso and the legs) with the same frequency of the swing, he can reach a status of self-sustained oscillation. When the forced oscillation is synchronized with natural oscillation, it is called resonance.

In this example, the restoring force was gravity. In an elastic system, such as a mass-spring oscillator, the restoring force resuslts from the compression and the elongation of the spring. This force is always in opposition with the displacement, thus returning (or restoring) the mass to equilibrium. Again inertia causes overshoot, thereby producing oscillation, and again we are in a self-sustained oscillation situation, or forced oscillation if energy sources and energy dissipative elements are included in the system.
For natural oscillation of a mass-spring system, the frequency is:

$$F_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{C.2}$$

where

- k is the stiffness of the spring;

- m is the mass;

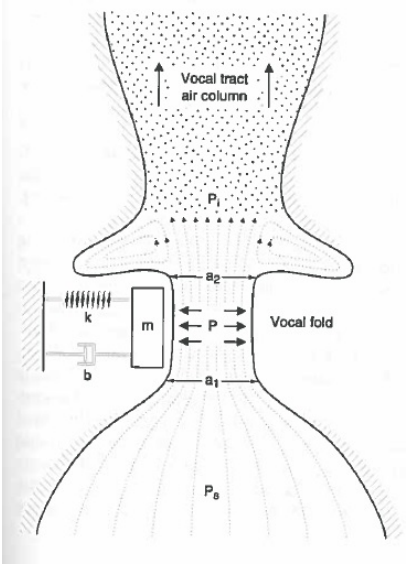


Fig. C.9 One mass model of the vocal folds, including airflow through the glottis, pressure against the tissue wall and an air column in the vocal tract.

A mass spring oscillator can be used to approximate the vocal fold. Stiffness relates to the elastic property of the tissues (the body and the skin). The elastic modulus, defined as a ratio of stress to strain for simple elongation, can be used to derive an effective stiffness of the vocal folds, but the procedure is quite difficult for complex deformations.
Consider each vocal fold to be a simple harmonic oscillator with mass m, stiffness k and an additional element called the *dampingcostantb* (Figure c.9). As applied to the vocal folds, this element represents the viscosity of the tissues, the energy absorber of the tissues. The stiffness k represents the effective stiffness of various tissue layers and m is the effective mass of the tissue in motion. The fluid pressure p in the glottis always acts perpendicularly to the tissue surface.
An expression of the *mean intraglottal pressure* over the medial surface of the folds has been derived on the basis of Bernoulli energy law. A simplified version of this expression is:

$$P = \left(1 - \frac{a_2}{a_1}\right)(P_s - P_i) + P_i \tag{C.3}$$

where

- a_1 is the cross-sectional area at the glottal entry;
- a_2 is the cross-sectional area at the glottal exit;
- P_s is the subglottal pressure;
- P_i is the input pressure to the vocal tract, or supraglottal pressure.

The quantity $(1-a_2/a_1)$ is a geometric factor that describes the shape of the glottis. The quantity (P_s-P_i) is called the transglottal pressure. In this case, where a single mass represents each fold, $a_1=a_2$ and the mean intraglottal pressure is simply the supraglottal pressure P_i. The transglottal pressure vanishes because $(1-a_2/a_1)$ goes to zero.

The resulting relation that the driving pressure P is equal to the supraglottal pressure P_i suggests that someting must happen above the glottis to change the pressure during the glottal cycle. The key element is the inertia of the air in the vocal tract: the delay in the response of this air column above the vocal folds causes another overshoot condition that aids in oscillation. This can be explained by considering an analogy between the moving air column and a moving solid mass. Pressure is analogous to force and a quantity called ineritance is analogous to mass. Ineritance for an air column is defined as

$$I=\frac{\rho L}{a} \tag{C.4}$$

where

- ρ is the density in the air column;
- L is the lenght of the air column;
- a is the cross-sectional area of the air column.

In according to the Newton law, one can write

$$P_i=I\frac{dU}{dt} \tag{C.5}$$

where

- U is the flow in the vocal tract, the volume velocity of the air column;
- $\frac{dU}{dt}$ is the the acceleration of the air column, the rate of change of flow.

According to this version of the Newton law, the vocal tract input pressure P_i is positive when the glottis is open and the acceleration is positive. This helps to drive the vocal folds outwards since the driving pressure P is equal to P_i. When the glottis is closed and the flowing is decreasing, the acceleration is negative, which makes P_i negative. This negative pressure, applied to the vocal folds, pulls them toghether. Thus, the folds are assisted in both opening and closing motions by supraglottal pressure. But this effect cannot by itself supply the necessary velocity-dependent force to transfer energy from the fluid to the tissue. Only in conjunction with vocal tract ineritance the proper asymmetrics between driving force and tissue velocity can be enstablished.

So, there are interactive mechanism between tissue velocity and supraglottal pressures. Both the two mechanism that can be used to achieve this syncronitìzation require delayed action between an upstream and a downstream event. When an air column is coupled to the oscillator, the buildup and collapse of forward momentum of the air column is delayed with respect to the glottal opening and closing. This delay creates asymmetry in supraglottal pressure with respect to tissue velocity, which drives the vocal folds in syncrony with their natural movement. Alternatively, when non-uniform tissue movement occurs, the delayed action between the upper and lower portions of the vocal folds creates an asymmetric driving pressure (vocal folds modes). Both of these delayed action mechanisms can work simoultaneously to improve the range of oscillation.

Consider the waveforms during sustained oscillation of the vocal folds illustrated in Figure c.10: c.10a shows vocal folds displacement x (solid line) at the centre of the glottis in the open portion of the cycle. The correspondig vocal folds velocity x is also shown (dashed line). Tissue velocity is positive during the outward movement and negative during the inward movement. In magnitude the tissue velocity is zero at the peak excursion and maximum before and after impact of the opposing vocal folds, but when direction is included the velocity is a decreasing function over the entire portion of a cycle.

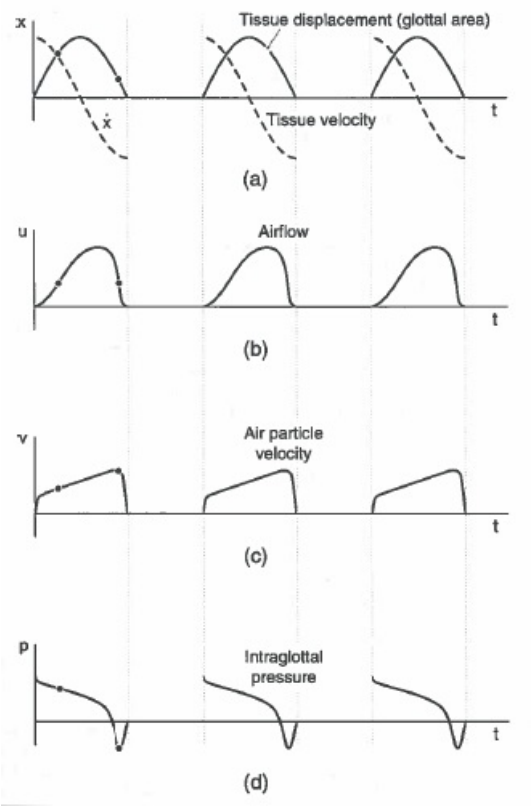


Fig. C.10 Waveforms during sustained oscillation: displacements and velocity of vocal fold tissue at the centre of the glottis (a), airflow (b), air particle velocity (c) and mean intraglottal air pressure (d) that drives the vocal fold.

Figure c.10b shows the glottal airflow U. It has been showed that a slow rise and an abrupt fall (that is the skewing of the waveform to the right) is a result of the inertia of the air column. The buildup of flow is delayed with respect to the movement of the folds (note that the peak in airflow comes later in time than the peak displacement). The decelartion of the flow is also delayed, but the fact that the glottal area is suddenly reduced to zero at the closure causes the sudden collapse of the flow. This asymmetry in the rising and falling of the flow is an important temporal feature.
Consider now the air particle velocity v in the glottis, which differs from the flow only for the cross-sectional area of the duct. Since the cross-sectional area of the glottis is proportional to lateral displacements of both folds, the shape of the air particle velocity waveform is obtained by dividing the flow waveform in figure c.10b by the displacement waveform in figure c.10a. This division is shown in figure c.10c. The important thing to note is that the air particle velocity is also asymmetric, displaying an average increasing to the right.

The waveform for the mean intraglottal pressure (and hence the driving force on the tissue) can be estimated on the basis of Bernoulli's energy law. In a duct, for constant flow, the following equation is valid:

$$P + \frac{1}{2}\rho v^2 = constant \tag{C.6}$$

Thus, for two identical flows in the opening and closing portion of the cycle (dots in figure c.10b) the glottal air pressure must be lower in the closing gesture than in the opening gesture: as air particle velocity increases from left to right, pressure in the glottis must decrease to keep to the left side of the equation c.6 constant.
It is important that the general downward trend of the mean intraglottal pressure matches the downward trend of the tissue velocity.

C.2.3 The Source filter theory of the vowels

Although voice production does not specifically deal with speech articulation, all human vocalizations require a particular configuration of the vocal tract. Unless the produced sounds are humming or hissing, the configuration is usually an open tract, as in a vowel: vowel formation is an integral part of the phonation process.
The acoustic properties of vowels have traditionally been described on the basis of

a source-filter theory: the sound source is the time-varying glottal airflow and the filter is the vocal tract; whereas the glottis produces a sound of many frequencies, the vocal tract amplifies, damps or suppress a subset of these frequencies for radiation from the mouth. This process is put into effect by the resonances that occur in the vocal tract; these resonances create standing waves inside the vocal tract, which interact with the original wave created by the glottis, originating interference. This interference is the main cause of the filtering effect of the vocal tract.

When an acoustic wave propagates in a tube, certain boundary effects have to be taken into consideration: the air particle velocity must go to zero at the tube wall is usually at its maximum at the center of the tube and decreases gradually toward the wall. The acoustic impendance would not be constant over the cross-section of the tube. To avoid multiple definition of a wave impendance over this cross-section, acoustic impendance and reflection coefficients must be reexamined on the basis of an average air particle velocity.
If the average particle velocity across the tube is multiplied by the cross-sectional area, an average flow in the tube is obtained. This average flow is conserved when the tube suddenly expands or contracts.

The acoustic impendance of the tube, with the wave traveling in only one direction and with no obstructions, can be defined as

$$Z = \frac{\rho c}{A} \left[kgs^{-1}m^{-4} \right] \tag{C.7}$$

where

- ρ is the air density;
- c is the sound velocity;
- A is the cross-sectional area of the tube;
- ρc is the free-space wave impendance.

The vocal tract can be approximated by a series of cylindrical tubes with varying diameters (figure c.11), which are used for ease of computation. They are arranged

in series to accomodate area changes in the vocal tract. With these cylindrical tubes, any complicated vocal tract shape can be modeled. Based on the new definition of acoustic impendance, a new reflection coefficient (at any two-tube interface) is defined as

$$r = \frac{\rho_2 c_2/A_2 - \rho_1 c_1/A_1}{\rho_2 c_2/A_2 + \rho_1 c_1/A_1} \tag{C.8}$$

where the subscripts 1 and 2 denote the acoustic properties of the adjacent tubes. In most speech applications, the air density and the sound velocity are costant throughout the airway. This means that $\rho_1 = \rho_2$ and $c_1 = c_2$ at any tube interface. The reflection coefficient can then be simplified in the form

$$r = \frac{A_1 - A_2}{1 + A_2} \tag{C.9}$$

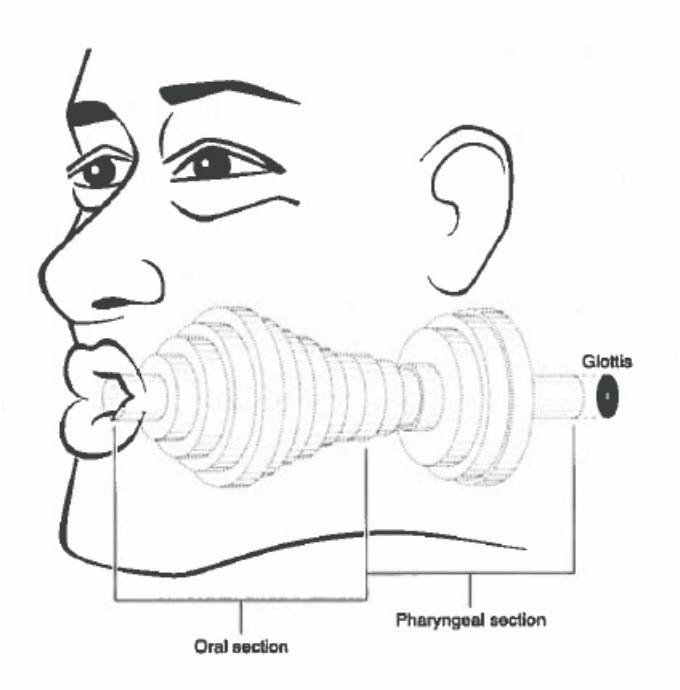


Fig. C.11 Cylindrical tubes approximation of the vocal tract for a simulated /u/ vowel.

In this formula, the convention is that the wave travels from tube 1 to tube 2. As A_2 approaches either zero or infinity, the reflection coefficients for closed-end or open-end tubes are obtained. For the closed-end configuration, $A_2 = 0$ and the reflection coefficient is $+1$: a positive compression with positive polarity. The

pressure at the closed end is twice the pressure of the incident wave. For an open-end configuration, $A_2 = \infty$ and the reflection coefficient is -1: again, complete reflection occurs, but the polarity is negative, and the sum of the incident and reflected pressure at the open end is zero (matching with atmospheric pressure outside the tube).

Vocal tract resonances

Resonance in a tube is the constructive interference of waves exeriencing multiple refelections; it is the essence of the vocal tract acoustics.
The human vocal tract, like the tube of a brass instrument, resonates at certain special frequencies produced by the sound source. These frequencies depends on the shape of the vocal tract, and determine many speech sounds from which syllabes and words are made.

The vocal human tract can be approximated with a tube of lenght L, closed at one end and open at the other. The glottis, the source, could be rappresented by a small hole in the closed end, which opens and closes itself periodically to supply acoustic energy.
When the compression/depression wave originated by the source reaches the end of the tube, suffers from a sudden expansion in the open space. This inverts the compression (polarity) and creates a new distrubance, a rarefaction of nearly the same magnitude. This rarefaction travels back down to the tube and becomes another incident wave as it approaches to the bottom, and there it is almost totally reflected. This reflection has the same polarity of the incindent wave, but reverse in respect of the source wave. Then, the new reflected wave travels to the top and change polarity again. Finally, after the second completed round trip, the pressure pattern is identical to the initial one.
A standing wave is created in this process. This pattern can be obtained if there is a relation between the period T of the wave and the the lenght of the tube L, and therefore the transit time t_0 for one round trip of the wave:

$$t_0 = (2n-1)(T/2) \tag{C.10}$$

The transit time may be expressed as the total distance divided by the speed of sound c:

$$t_0 = \frac{2L}{c} \tag{C.11}$$

and so we can write

$$\frac{2L}{c} = (2n-1)(T/2) \tag{C.12}$$

By considering the frequency F of the source as the inverse of the period T, the last equation can be rewritten as

$$F = (2n-1)\frac{c}{4L} \tag{C.13}$$

A so called formant is a resonances of the vocal tract; with this definition, the equation c.13 identifies the formant frequencies of a closed-open uniform tube. Formants are the normal modes of the vocal tract air column.

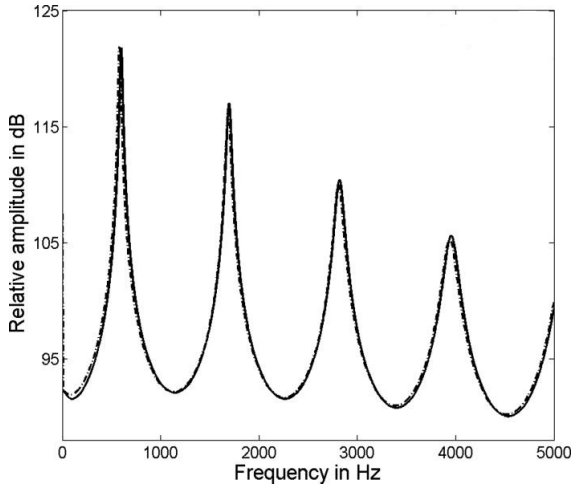


Fig. C.12 EGG spectra of a vocalization.

To distinguish between each of the formant frequencies, a subscript n is placed. The formant frequencies can be written as

$$F_n = (2n-1)\frac{c}{4L} \tag{C.14}$$

If we assume an average man vocal tract length of 17.5 cm and a sound velocity of 35000 cm/s, the formant frequencies are

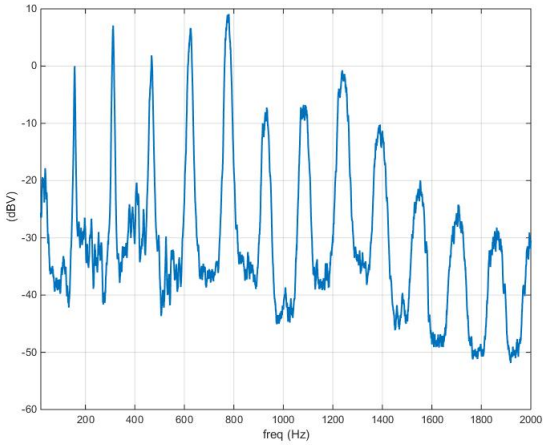


Fig. C.13 Acoustic spectra of the /a/ vowel.

$$F_n = (2n-1)\frac{35000cm/s}{70cm} = (2n-1)500Hz \tag{C.15}$$

The spectrum of these frequencies are a line spectrum, meaning that the tubes resonates at precise frequencies only. Considering that some acoustic energy is radiated from the open end of the tube, some is lost through the glottis into the lungs and som is absorbed by the tissues and by the friction of the air particles, the tube becomes less selectrive in its response to frequency, and the peaks becomes thicker. The source of phonation, the glottis, produces a pressure wave with a spectra like the one reported in figure c.12 (EGG). This spectra is filtered by the vocal tract, which amplifies some frequencies (according to its resonances) and attenuates some others. The results is the voice spectra, which is presented in figure c.13.

C.3 References

1 Ingo R. Titze, Principles of voice production, National Center for Voice and Speech, Denver Colorado, 2000.

2 Kosztyła-Hojna B., Rogowski M., Usefulness of video-laryngo-stroboscopy in the diagnosis of laryngeal pathology, Pol Merkur Lekarski, Vol 14(83), pp 413-416, May 2003.